

# THIS WEEK

## EDITORIALS

**WORLD VIEW** How Hollywood has got it right when it comes to stammering **p.7**

**WASPS** Beat of a drum turns queens to workers **p.9**

**EGYPT** Citizens act to protect Cairo antiquities museum **p.10**



## Earth 2.0

*The hunt is on for a distant planet similar to our own. Astronomers should decide just how similar it needs to be, before the candidates start pouring in.*

**T**he search for a second Earth has long enthralled readers of science fiction. What rich and varied life could it contain? What would such a discovery mean for humanity's own place in the Universe? How many similar planets are out there? The question is more than a philosophical puzzle, and it comes with a hard scientific edge that should be considered sooner rather than later. As the search for planets beyond the Solar System widens and public interest in the quest grows, at which point should astronomers declare the hunt for another Earth a success?

Hundreds of candidate planets have been identified, and some have been profiled, if not as a second Earth, then as signs that the search is heading in the right direction. Last month, NASA announced the discovery of the smallest extrasolar planet yet: Kepler-10b, which has 1.4 times the diameter and 4.6 times the mass of Earth, and was discovered by NASA's Kepler spacecraft (see page 24). Although the planet orbits too close to its star to support life, the news was heralded by some media outlets as a landmark in the search for a new Earth, particularly because Kepler-10b is the first exoplanet with a dense and rocky core.

Attention on Kepler's mission will intensify again this week, as NASA publicly releases a batch of its data (see page 53). The satellite focuses on a single point in the sky, where it can keep track of some 150,000 stars. Kepler observes the decrease in the brightness of these stars as planets pass in front of, or 'transit', them, and the findings are used to target telescopes on the ground.

It takes three to four such confirmed transits before astronomers are confident that they have found a planet, which makes it too soon to be sure whether Kepler has found a world truly similar to Earth. (By definition, Earth-like planets orbiting a star similar to the Sun pass in front of their stars about once a year, and Kepler has only been in place for about 18 months). All exoplanets confirmed to date orbit much closer to their stars than does Earth; they are too close for conditions to allow the existence of liquid water, which is what defines a star's 'habitable zone'.

As more data are analysed, they will probably produce a string of reports of ever-smaller planets, until we get an Earth-sized example. Many of these small planets are likely to orbit M-dwarfs, by far the most numerous type of star in the Universe (see page 27). The habitable zone around these stars is very narrow, but Kepler may find a rocky planet there. Would that be the first Earth-like planet? Probably not if, as seems likely, it were to be tidally locked, so that one side faced permanently towards the star.

What about planets that orbit larger stars? Does a first Earth-like planet have to orbit in the habitable zone of a G2-type star, similar to the Sun? If so, must the planet be Earth-sized? And is the focus on a habitable zone defined in terms of liquid water appropriate? As the Universe reveals its secrets, we discover it to be a more diverse and stranger place than we had anticipated. Would it be so odd to conceive of life on a dry or frozen world? Must the first Earth-like

planet be capable of supporting life, or human life in particular?

The answers to these questions are important because the public-relations rewards of planet-hunting — and planet-finding — are great. The temptation to hype each discovery is equally large, but so is the scope for confusion and public scorn, especially given the rabid response on some blogs to NASA announcements. Set the bar for 'Earth-like' planets too low, and a string of repeated discoveries could be overwhelming. Set the bar too high, and a planet that meets the strict criteria may not emerge at all. If that were to happen, the Kepler mission would risk being viewed as a failure — which it most certainly is not.

Amid the excitement of exploring a new frontier, astronomers should pause to consider the public reaction to their work. Then they should decide how a standard should be set. Perhaps a reasonable starting point would be to define an Earth-like planet as one of similar size to Earth, orbiting in the habitable zone of any star, and not tidally locked. More important than the details of the definition is that the relevant criteria are established before the claims start to pile up. To announce the discovery of the first Earth-like planet would be a stunning success. To announce it more than once could look like carelessness. ■

***"As more data are analysed, they will produce a string of reports of planets."***

## Preserve the past

*Historic scientific collections deserve better than to gather dust.*

**W**hen the celebrated anatomist Antonio Scarpa died in 1832, he left an extensive collection of anatomical preparations to his university in Pavia, Italy. The collection includes his own head which, pickled, now presides grimly over his legacy in a protected museum.

Across Europe, a distressingly high number of historic scientific collections — from herbaria to minerals — are being lost or left to rot in universities. As many are autonomous institutions, they can't be told what to do by governments, they are mostly poorly funded and they tend to be oriented to the future, not the past. Historic collections have to compete for space and resources with active researchers, and are rarely prioritized.

Germany may have come up with a way to break out of this dilemma. Earlier this week, the Wissenschaftsrat, the nation's influential science council, issued a detailed list of recommendations that declares that scientific collections of potential research value should be handled as

research infrastructures. It says that universities have a duty to preserve collections that it describes as being 'in critical condition' and make them available to internal and external researchers — and to integrate them as appropriate into teaching programmes.

The council also details how this should be done. Universities, it says, together with Germany's research museums and the country's main granting agency, the DFG, should develop criteria to assess the scientific merit of a collection. These criteria should then be applied in a hard-nosed fashion so that inferior collections are closed or transferred elsewhere. Furthermore, historic scientific collections in universities should be allocated the space they need, including a room for researchers to work on them and for exhibitions.

This prioritization is important. Historic objects frequently turn out to have great — often unexpected — value for cutting-edge research. Well-preserved old bones, for example, are a treasure trove for modern palaeontologists wielding new DNA-based analytical technologies. Old herbaria can similarly feed the curiosity of today's plant geneticists. Historic collections can also be unique resources for social scientists, particularly science historians. So the Wissenschaftsrat's endorsement of their fundamental value is extremely welcome.

But will the recommendations be taken up? Most probably yes. The Wissenschaftsrat has serious clout because it comprises representatives from both the federal and state governments, as well as scientists. Its procedures are systematic and its analyses are thorough. It makes no recommendations that its members know they will not be able to pay for. It could, however, take some time for the recommendations to be implemented.

In the case of scientific collections, the Wissenschaftsrat proposes that the federal government issues a call for proposals for a five-year project to coordinate the efforts of universities to save their collections.

Scientists at German universities — which are funded by state governments, and constitutionally banned from receiving infrastructure funding from the federal government — should find short-term grants from research agencies or foundations to upgrade, restore and make their collections available. The universities themselves should then

provide overheads for ensuring that the collections are looked after in the long-term, a sum that the Wissenschaftsrat says should be modest.

These recommendations did not emerge from a vacuum. In 2004, the DFG supported a five-year project to identify and catalogue

collections in German universities. It identified more than a thousand, of which nearly 300 were shown to have been lost or destroyed. Herbaria that were once state-of-the-art, for example, were confined to dusty cellars or stuffy attics when classical botanics fell out of research fashion.

This DFG programme won the admiration of scientists in many countries that still have no national catalogue of the treasures hidden in their own universities and no systematic way of preparing one. The Wissenschaftsrat's initiative is now winning admiration for its long-term vision and political commitment.

There will certainly be battles to come. Simply saying that space should be made available for collections isn't helpful if there is genuinely no space to be had, for example. But the value that the Wissenschaftsrat now places on collections should make such battles easier to win. Research organizations in other countries should look to see if they could follow its lead. A collection deemed scientifically valuable doesn't need to be as peculiar as Scarpa's head to make it worth preserving, but it needs the same protections and accessibility. ■

# Tough on truth

*The Global Fund should be praised for coming clean about fraud by grant recipients.*

**"F**raud plagues global health fund," screamed the title of an article published last month by the Associated Press, which alleged that: "A \$21.7 billion development fund backed by celebrities and hailed as an alternative to the bureaucracy of the United Nations sees as much as two-thirds of some grants eaten up by corruption, The Associated Press has learned."

Journalistic scrutiny of aid is welcome and revelations of widespread and large-scale fraud by recipients of grants from the Global Fund to Fight AIDS, Tuberculosis and Malaria would be a big deal. The fund, created by the highly industrialized countries of the G8 forum in 2002, now accounts for one-quarter of all international financing to fight AIDS, two-thirds of that for tuberculosis, and three-quarters of that for malaria. But despite using the phrase "has learned" — journalist shorthand for a scoop — the Associated Press (AP) article's central claims contained no new revelations. The frauds mentioned — involving grants to Mali, Mauritania, Djibouti and Zambia — had already been made public by the fund itself.

The sums involved in the reported fraud cases amount to US\$39 million, of \$13 billion that the fund has disbursed, but other fraud cases have no doubt so far gone undetected. Although any corruption is too much, to keep it down to these levels would be an achievement, given the realities of putting large amounts of money into any country or project, not least those where corruption can be rife.

Nonetheless, Sweden, Germany and Ireland have responded with suggestions they may suspend their pledges to the fund for the period covering 2011–13. As fund members they are well aware of how it handles

corruption, so their response is probably partly a reaction to the wide publicity that the AP article received in the international media, and the sensationalist and exaggerated claims about the scale of the problem — no government, accountable as it is to taxpayers, wants to be seen as lax on corruption. As *Nature* went to press, reports suggested that funding from these countries would be restored, while Sweden has also since said that it is happy with the way the Global Fund is dealing with the problem.

The reputation of the fund — which by its own estimates saved more than 4.9 million lives by 2009 — has been unfairly tarnished, and its fund-raising efforts perhaps hampered at a time when the economic crisis is already making donors reconsider the size of their contributions.

When it comes to being transparent over problems of corruption in recipient countries the Global Fund has been far better than most aid donors or agencies. It has openly tackled corruption — with a 'zero tolerance' policy, suspending grants at the first whiff of wrong-doing, and working with recipient countries to bring fraudsters to justice and recover what misdirected money it can. Could it do more? Yes: for example, by strengthening oversight further. But it is already well down the road to effectively tackling corruption.

The same cannot be said for many of the alphabet-soup of aid agencies, which choose not to publicise their own uncovered fraud cases, perhaps out of fear of damaging their image, and losing donors. Several observers have been quick to point out that if the AP article has an upside, it is to have drawn renewed attention to fraudulent use of funds by such agencies. The fight against aid corruption has generally improved markedly since the 1990s, but many agencies still fall far below the high bar set by the Global Fund. Meanwhile, astonishingly, the fund's own fraud investigations have been hampered because the

United Nations Development Programme, which manages some of its grants, has refused to allow the fund access to its records. Scrutiny should be welcomed, but honesty should not carry so high a price. ■

➤ **NATURE.COM**

To comment online,  
click on Editorials at:  
[go.nature.com/xhunqy](http://go.nature.com/xhunqy)

A. LEWIS



## Listen to the lessons of *The King's Speech*

A film that shows King George VI struggling with a stammer could raise awareness and change treatments, says **Peter Howell**.

**T**he recently released film *The King's Speech* looks set to sweep the board at the Oscars later this month, and actor Colin Firth has been widely praised for his depiction of King George VI as he seeks to overcome a debilitating stammer. As a scientist who studies stammering (or stuttering as it is also called), I liked the film. Many people still poke fun at those who stammer, and find their difficulties amusing. Perhaps the film's success, with raised public awareness of the condition, will help to address such prejudice.

We cannot be sure of the accuracy of the historical events the film portrays, although many of the scenes seemed to me to ring true. In many ways, we understand a lot more than we did in George VI's day. In others, we are just as blind. Although 1 in 20 children stammer, most grow out of it. Of the 1 in 100 people who still stammer as they enter their teenage years — including, it seems, George VI — few go on to recover. For them, it becomes a question of how to manage the condition. This is an important lesson from the film: we should not give false hope to the parents of children who are unlikely to recover from their stammer.

Techniques to control stammer in adults can be effective, as the film shows. But without spoiling it too much for those who have not seen it (and you should!), it is fair to say that getting people who stammer to repeatedly shout rude words would no longer be suggested as a technique, if it ever was.

Those who stammer, however, do often find they can speak fluently when they are prevented from hearing their own voice. The film illustrates this when George VI is shown speaking fluently when music is played. Playing their speech back to people in real time, but delayed or shifted in frequency, can also help. We still do not fully understand why this happens, but prosthetic devices are available that produce this effect and bring about improvement. The most discreet of these cost up to US\$5,000, although opinion in my community is divided as to whether such technology is appropriate, because the stammering returns when the device is switched off.

More controversially, some researchers believe they can induce fluency in people, children in particular, using verbal operant procedures, similar to the reward and punishment techniques used to train dogs and other animals. The most common of these is the Lidcombe Program, developed by researchers at the Australian Stuttering Research Centre at the University of Sydney. Although some data suggest that the technique works, the numbers are too low to draw firm conclusions. A major sticking point is that there is no commonly agreed method to diagnose the disorder in childhood. The Australian group considers children who

commonly repeat whole words as stammerers. In my view, this is wrong and could skew the results.

Much of the controversy over treatments such as the Lidcombe Program and prosthetic devices centres on whether the benefits linger when the treatment ends. Such carry-over would demand that stammering is at its heart a learned behaviour that can be undone. I do not think it is.

So, what causes people to stammer? And how can we distinguish between children who will recover from their stammer, and those who will not. These are questions of great importance for both sufferer and society. Comparison of people who do or do not recover suggests that several factors are important. Biological (genetics and brain differences), linguistic and motor factors, and type of stammering

symptom are reliably reported to differ between such groups.

Much attention in the media has focused on a 'stuttering gene' after the discovery of genetic mutations in members of a consanguineous (inter-marrying) Pakistani family. The gene identified codes for proteins involved in cellular lysosome function, which removes damaged molecules and viruses. Other geneticists, including Simon Fisher, director of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, point out that the lysosome function is broad and more research is needed on how it could affect the central nervous system of people who stammer.

A study on a Chinese family suggests a more plausible genetic basis, as it reported mutations in genes that seem to affect parts of the brain (the basal ganglia) previously implicated in the disorder.

My work has shown that the severity of stammering symptoms in eight-year-old children can be used to predict whether the children are likely to still stammer as teenagers and adults. Wider discussion is needed about how this information should be used by therapists — for example, whether to intervene with treatment. As most people will recover in time, watchful monitoring of children for signs that their stammering is worsening may be a better approach.

If Colin Firth's performance is a fair reflection of George VI's speech, then the types of stammer he produced — prolonged consonants and repetition of the first part of words — would, as a child, have suggested he had little chance of recovery. We should make more effort to pass this information on to parents. Expectations must be realistic, but, as *The King's Speech* shows, they can still be optimistic. ■

**Peter Howell** is professor of experimental psychology at University College London and author of *Recovery from Stuttering*.  
e-mail: p.howell@ucl.ac.uk

**MUCH OF THE  
CONTROVERSY  
CENTRES ON  
WHETHER THE  
BENEFITS LINGER  
WHEN THE  
TREATMENT  
ENDS.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/nad2nl](http://go.nature.com/nad2nl)



# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## CANCER

### Weakening tumour defences

Blocking just one enzyme could make some tumours easier to kill with radiation therapy.

Fei-Fei Liu at the University of Toronto in Ontario, Canada, and her team found that head and neck cancer cells treated with small RNA molecules that silence a gene called *UROD* are unusually susceptible to ionizing radiation. *UROD* encodes an enzyme involved in producing iron-containing haem molecules, and reducing its levels caused oxidative damage and cell death. When implanted into mice that were then treated with radiation, tumour cells with lowered levels of the enzyme grew more slowly than cells containing normal levels.

Furthermore, head and neck tumour samples expressing the lowest levels of *UROD* tended to come from patients who responded well to treatment. Drugs that inhibit *UROD* could make radiation treatments more effective, the authors say. *Sci. Transl. Med.* 3, 67ra7 (2011)

## METAGENOMICS

### Cow spills guts for biofuels

A dearth of ways to efficiently digest the plant fibre cellulose has stymied efforts to develop plant-based biofuels. Genomic

sequencing of bacteria from the cow's digestive system has turned up enzymes that might address this problem, reports a team led by Edward Rubin of the Lawrence Berkeley National Laboratory in Berkeley, California.

The researchers sequenced 268 gigabases of DNA from microbes that adhered to plant material in the rumen of the cow, a champion cellulose digester. The microbes were collected directly from the cow rumen through a port (pictured). The researchers identified 27,755 genes possibly involved in carbohydrate digestion, and produced protein from 90 that had similarities to known

enzymes with particular attributes. More than half demonstrated enzymatic activity against a panel of 10 cellulose-containing plants. *Science* 331, 463–467 (2011)

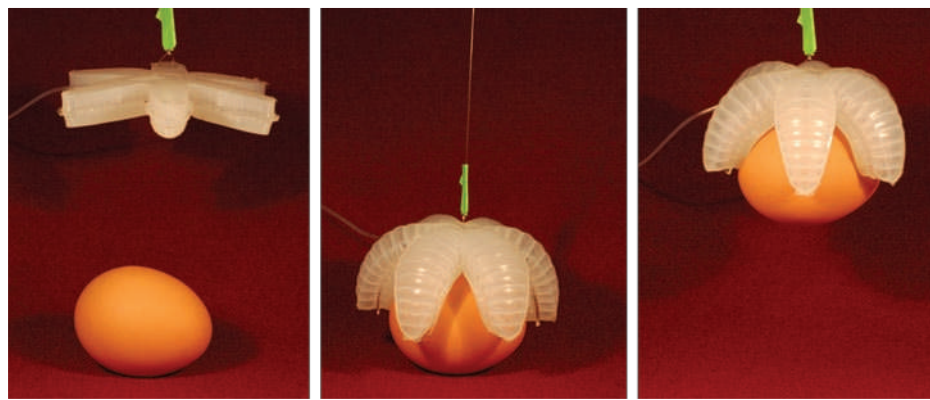
## NEUROSCIENCE

### Omega-3s affect brain signalling

A diet low in omega-3 fatty acids — typically found in fish oils — has been associated with mood disorders. To find a molecular link, Sophie Layé at the University of Bordeaux in France, Olivier Manzoni at the French National Institute of Health and Medical Research in Marseilles and their co-workers fed mice a diet that was either high or low in omega-3s. They then looked

at the ability of neurons in the prefrontal cortex, a brain region thought to mediate emotional behaviour, to alter the strength of their connections — a process known as synaptic plasticity. The authors focused on lipid signalling molecules called endocannabinoids and their receptors, which are involved in this process.

They found that mice whose diets were low in omega-3s had lower levels of the fats, and reduced synaptic plasticity, in the prefrontal cortex. This was due to the decoupling of the cannabinoid receptors from certain proteins that normally bind to them. Mice fed the low-omega-3 diet also showed behavioural signs of depression and anxiety. *Nature Neurosci.* doi:10.1038/nn.2736 (2011)



## MATERIALS

### 'Soft' robot has deft touch

Robots made from hard materials are not well equipped to handle fragile objects, so researchers have created prototypical 'soft' robots from elastic polymers. These can perform delicate tasks such as picking up eggs.

George Whitesides and his colleagues at Harvard University in Cambridge, Massachusetts, embedded balloon-like channels in moulded silicones to create

materials that change shape in specific ways when air is pumped into the channels. This can generate complex motion from a single source of pressure, leading to designs such as the team's starfish-shaped gripper (pictured).

With appropriate materials, this technology might eventually produce robots that can handle heavy loads or conduct electricity. *Angew. Chem. Int. Edn* doi:10.1002/anie.201006464 (2011)



F. ILIEVSKI

J. L. GUERSTAD



## POPULATION BIOLOGY

## Whales found where whaling was

The North Atlantic right whale (*Eubalaena glacialis*) is one of the world's rarest cetaceans, and little is known about its wintering or summering grounds, hampering conservation efforts. Now, researchers have documented signs of the whale in an area that was a whale-hunting ground in the late 1800s.

David Mellinger at Oregon State University in Newport and his colleagues carried out a year-long acoustic survey at five sites in and around the Cape Farewell Ground waters, an area about 500 kilometres east of southern Greenland. The team recorded more than 2,000 whale communication calls, mainly between July and November, suggesting that the area is still an important summer ground for the creatures.

The data will help to guide the relocation of shipping lanes and restrictions on vessel speed to prevent collisions with the animals.

*Biol. Lett.* doi:10.1098/rsbl.2010.1191 (2011)

## ECOLOGY

## Kill one species to save the rest

The loss of one species from an ecosystem can have unpredictable — and on occasion catastrophic — cascading effects. A modelling study suggests a strategy for rescuing a troubled ecosystem: selectively remove one or more additional species.

Sagar Sahasrabudhe and Adilson Motter of Northwestern University in Evanston, Illinois, showed that removing or partially suppressing one or more species in a food web at key time points after one member has become extinct saves other members of the web from the same fate. The duo used several model food webs, as well as two webs modelled with data

derived from real ecosystems — the Chesapeake Bay off Maryland and Virginia, and the Coachella Valley in Southern California.

The idea — a controversial one that may not sit well with some conservationists — relies on the fact that ecosystem networks tend to shift to a different stable arrangement after losing members.

*Nature Commun.* doi:10.1038/ncomms1163 (2011)

For a longer story on this research, see [go.nature.com/2sckoo](http://go.nature.com/2sckoo)

## PHYSICS

## Laser travels forwards and back

One way to identify poisonous gases, or the vapours released by explosives, is to detect the effect of these molecules on laser light beamed through them. But practical detection devices should send and collect the laser beam from the same side of the gas cloud. Arthur Dogariu and his colleagues at Princeton University in New Jersey have taken a step towards this goal by demonstrating backwards lasing in air.

They used an ultraviolet 'pump' laser to break up oxygen molecules. The same laser then excited the molecular fragments into generating an infrared beam. The region in the air that the pump laser was focused on was about 100 times longer than it was wide. So half of the infrared light was emitted forwards, and the other half travelled backwards towards the source. The returning beam carried fingerprints of other molecules in the air.

*Science* 331, 442–445 (2011)

## ANIMAL BEHAVIOUR

## Drummed into submission

Paper wasps divide the work of the colony between different castes: workers build and defend the nest, whereas individuals destined to become queens lay eggs. Wasps do not inherit these

## COMMUNITY CHOICE

The most viewed papers in science

## MATERIALS SCIENCE

## Lights on for drug delivery

**HIGHLY READ**  
on [pubs.acs.org](http://pubs.acs.org)  
in Dec 2010

A biocompatible gel that sheds its load when exposed to ultraviolet light might be used for controlled delivery of drugs and other molecules inside the body.

Weihong Tan at the University of Florida in Gainesville, Xiaoling Zhang at the Beijing Institute of Technology and their colleagues created this gel by first dissolving polymers decorated with DNA strands in water. They added DNA fragments that bind to the polymer-bound DNA, crosslinking the polymers to form a gel. These fragments also carry light-sensitive azobenzene molecules that, if hit by ultraviolet light, cause the crosslinks to break, releasing any trapped molecules. The gel unloaded a variety of test cargoes, including nanoparticles, an enzyme and the cancer drug doxorubicin, in a matter of minutes. Visible light restores the hydrogel's structure. *Langmuir* 27, 399–408 (2010)

roles, but are instead set on a particular developmental path during the larval stage. Researchers have discovered how this occurs in the genus *Polistes*: queens use their antennae to drum near to or on nest cells containing larvae to turn them into workers.

Sainath Suryanarayanan at the University of Wisconsin–Madison and his group used an electrical device to simulate this drumming on colonies that produce *Polistes fuscatus* wasps (pictured) destined to become queens. The wasps that emerged had the lean body type of workers. The link between the drumming — which for larger *Polistes* species is audible outside the nest — and gene-expression changes is not clear. *Curr. Biol.* doi:10.1016/j.cub.2011.01.003 (2011)



Work in mice suggests that chemical modifications to the DNA may explain the difference.

Shusaku Uchida and Yoshifumi Watanabe at Yamaguchi University in Japan and their colleagues subjected two genetically distinct strains of mice to chronic stress and then measured various proteins involved in neuronal growth and maintenance. The strain known to succumb to stress had lower than normal levels of a protein called GDNF in the brain's striatum. The resilient strain had higher amounts.

The team found that histones — proteins that package up DNA and regulate gene transcription — on a section of the *Gdnf* gene were modified differently between the two strains. This led to *Gdnf* repression in the susceptible mice and increased expression in the more resilient strain. *Neuron* 69, 359–372 (2011)

## NEUROSCIENCE

## Root of resilience under stress

Some individuals react coolly to stressful events, whereas others slip into depression.

➔ **NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

E. RESCHKE/PHOTOLIBRARY

# SEVEN DAYS

The news in brief

## POLICY

### GM alfalfa ruling

A long-running battle over genetically modified (GM) alfalfa has ended with the US Department of Agriculture (USDA) deciding that farmers can plant the crop without restriction. Just a month ago, the USDA put forth a draft plan that would have limited where the herbicide-resistant crop could be planted, addressing organic farmers' concerns about contamination of their fields. But the plan came under fire from politicians and lobbyists, and on 27 January, the USDA reversed its stance.

### Polio pledges

Billionaire Bill Gates has called for a final push to eradicate polio. The Bill & Melinda Gates Foundation will add US\$102 million to its annual \$200-million pledge for the cause, Gates announced on 28 January. The United Arab Emirates and the United Kingdom also made funding commitments last week. The World Health Organization's Global Polio

## SOUND BITE

“It’s like a math teacher not believing in algebra.”

William Wallace, Washington DC representative of the National Association of Biology Teachers in McLean, Virginia, responds to a study suggesting that most biology teachers in publicly funded high schools are uncomfortable with teaching evolution. See [go.nature.com/yfgex8](http://go.nature.com/yfgex8) for more.



AMR NABIL/AP

## Cairo museum survives Egyptian looters

Amid the public unrest in Egypt last week, Cairo's world-famous Museum of Egyptian Antiquities seems to have been spared serious damage. Looters tried to steal two mummy skulls and damaged some 100 items at the museum on 28 January. But other citizens stepped in to guard the museum, which holds precious antiquities including Tutankhamun's

death mask, before the army secured the building the next day (pictured). Zahi Hawass, head of Egypt's Supreme Council of Antiquities in Cairo, says that the damaged items can be restored, adding that the looters mainly stole jewellery from the gift shop. Archaeological sites outside the city may not have been so lucky, reports suggested as *Nature* went to press.

Eradication Initiative, which vaccinates children against the virus, needs \$720 million to fill a funding gap in 2011–12. Polio remains endemic in Afghanistan, Pakistan, India and Nigeria.

### US budget tension

In his annual State of the Union address on 25 January, US President Barack Obama stressed the need to invest in research — particularly in biomedical science and clean-energy technology — but he also proposed a freeze on annual domestic spending for the next five years. Obama's 2012 budget request is expected in mid-February. Congress has yet to vote on the budget for 2011, which is expected to include tough austerity measures.

### Weak innovation

Member states in the European Union are losing their lead in innovation over Brazil and China, and are not catching up with the United States and Japan, according to the European Commission's annual *Innovation Union Scoreboard*, released on 1 February. The report, based mainly on data from 2008 and 2009, says that Europe's patent revenues and business research spending are particularly poor compared with those of Japan.

### China's wind win

China has now installed more wind-power capacity than has been deployed by the United States. The American Wind Energy Association, in a 24 January report, said that its wind industry installed

5.1 gigawatts of capacity last year — half the 2009 total — to reach a total capacity of 40.2 gigawatts. Data released on 12 January by the Chinese Renewable Energy Industries Association show that China installed 16 gigawatts in 2010 to reach 41.8 gigawatts.

## RESEARCH

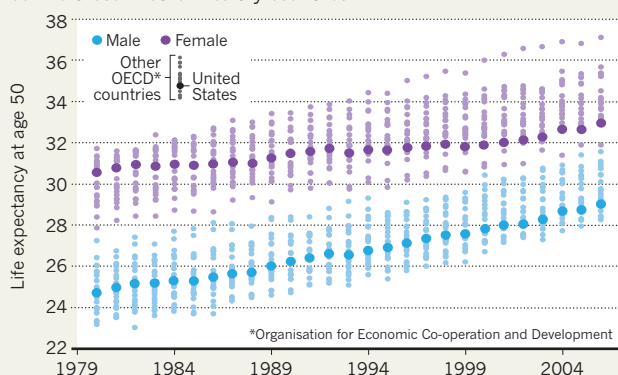
### Vostok drilling

Russian researchers drilling down to the sub-glacial Lake Vostok, 3,750 metres under Antarctica's ice sheet, told *Nature* that they hope to reach within 20 metres of the lake's surface by 6 February — the last day of Antarctic summer operations. But they do not expect to penetrate the pristine lake this season. Operations will resume in December.

SOURCE: NRC

## THE RACE TO LONGER LIVING

Improvements in life expectancy in the United States are lagging behind those in other wealthy countries.



## Smoking gun

Life expectancy in the United States is lower than in many other wealthy countries (see chart), despite the country's huge health-care spending. The popularity of smoking in past decades and today's rising obesity rates are partly to blame, a report from the US National Research Council concluded last week.

## LHC extension

The Large Hadron Collider (LHC) will run until the end of 2012. The particle accelerator at CERN, Europe's high-energy physics laboratory near Geneva, Switzerland, had been due to stop in 2011 for a year-long shutdown to upgrade its collision energy to 14 teraelectronvolts (TeV). But after a meeting in

Chamonix, France, last week, scientists decided to give themselves an extra year to collide particles — currently at 7 TeV but potentially reaching 8 TeV — with the hope of collecting enough data to spot the Higgs boson. See [go.nature.com/wd9fug](http://go.nature.com/wd9fug) for more.

## Dengue control

Six thousand mosquitoes genetically engineered to be sterile were released in Pahang state, Malaysia, on 21 December. But the field trial, which aims to control dengue fever by suppressing mosquito populations, was announced only on 26 January, by Malaysia's Institute for Medical Research in Kuala Lumpur. Scientists and advocacy groups were surprised, as they believed the trial had been postponed.

It follows larger trials on the Caribbean island of Grand Cayman in 2009 and 2010, all run by Oxitec, a company based in Oxford, UK.

## Intel research

Intel is investing US\$100 million in US universities over five years, by opening six to eight science and technology centres on university campuses throughout 2011. Announcing the investment on 26 January, the company said that its first centre will be at Stanford University in California, researching visual computing.

## Solar sailing

Two trials of solar sails, which use the pressure of photons from the Sun to propel spacecraft, are going well. On 26 January, Japan's Aerospace Exploration Agency said that its Ikaros space capsule, which has a 200-square-metre solar sail, had completed six months of space flight and would have its mission extended to March 2012. Six days earlier, NASA deployed a 10-square-metre solar sail, NanoSail-D, in low-Earth orbit.

## BUSINESS

## Pharma pruning

In the latest round of drug-industry cutbacks, Abbott Laboratories, headquartered in Abbott Park, Illinois, will

## COMING UP

### 4 FEBRUARY

The European Council discusses energy policy and European innovation, at a summit meeting in Brussels. [go.nature.com/olubai](http://go.nature.com/olubai)

### 10–11 FEBRUARY

The future of the delayed LISA Pathfinder mission (see *Nature* **469**, 280; 2011) will be assessed by the European Space Agency's Science Programme Committee.

cut 1,900 jobs, of which the firm said "a small number" were in research and development. Elan, based in Dublin, will cut 130 jobs — about 10% of its workforce. Around half of these are scientists, and most are based at the biotech firm's research and development facility in South San Francisco.

## Biotech buy

Biotechnology giant Amgen will pay US\$425 million to acquire a cancer-vaccine company. Amgen, based in Thousand Oaks, California, could also spend up to \$575 million in milestone payments for BioVex Group. The biotech, based in Woburn, Massachusetts, develops tumour-killing viruses that also provoke immune responses against the cancer.

## Drug deal near

Drug-maker Sanofi-aventis of Paris was reported to be nearing a deal to buy Genzyme of Cambridge, Massachusetts, as *Nature* went to press. Sanofi is said to have boosted its rejected US\$18.5-billion bid. Last month, Sanofi suffered a setback when late-stage clinical trials showed that its anticancer drug iniparib failed to slow advanced breast cancer.

**NATURE.COM**

For daily news updates see: [www.nature.com/news](http://www.nature.com/news)

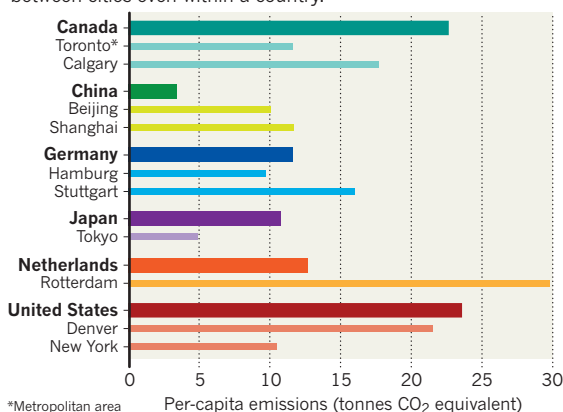
SOURCE: ENVIRONMENT AND URBANIZATION

## TREND WATCH

Although cities generate most of the world's greenhouse-gas emissions, their per-capita emissions vary widely (see chart). New York, for example, has half the per-capita emissions of Denver, thanks to its denser population and lower reliance on cars, notes a study published last month (D. Hoornweg *et al. Environ. Urban.* doi:10.1177/0956247810392270; 2011). Rotterdam's per-capita emissions are particularly high because its port attracts industry and fuelling of ships.

## CITIES' CARBON EMISSIONS

Average per-capita greenhouse-gas emissions vary widely between cities even within a country.





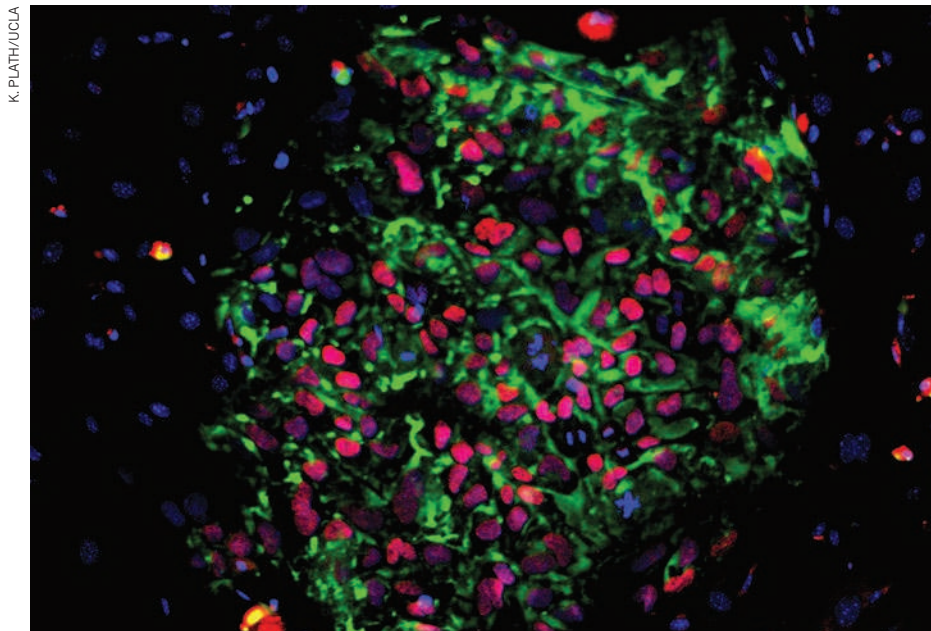
# NEWS IN FOCUS

**ASTRONOMY** US bidders cry foul after radio telescope heads to the far east **p.14**

**CHINA** A superpower sets bold plans to boost research and innovation **p.15**

**SOCIAL SCIENCE** Researchers choose their top ten questions **p.18**

**EXOPLANETS** The hunt is on for worlds that are like our own **p.24**



K. PLATH/UCLA

Induced pluripotent stem cells retain 'memories' of the adult cells from which they are derived.

## MEDICINE

# Flaw in induced-stem-cell model

*Adult cells do not fully convert to embryonic-like state.*

BY ELIE DOLGIN

Medical researchers' hopes of replacing politically fraught embryonic stem (ES) cells with stem cells derived from adult tissues have suffered a setback. Induced pluripotent stem (iPS) cells, created by turning back the developmental clock on adult tissues, and ES cells display similar gene-expression patterns, and both can produce any of the various tissues in the human body. But patterns of epigenetic changes — alterations that affect gene expression without changing the DNA sequence — tell a different story about iPS cells, a team led by Joseph Ecker, a molecular geneticist at the Salk Institute in La Jolla, California, reports online in *Nature* this week<sup>1</sup>.

"They are slightly different creatures," says Chad Cowan, a stem-cell biologist at Massachusetts General Hospital in Boston who was not involved in the work. The finding suggests that iPS cells may not be suitable substitutes for ES cells in modelling or treating disease.

Ecker and his colleagues analysed patterns of DNA methylation, a type of epigenetic change, across the genomes of 15 cell lines. These included four human ES cell lines, five iPS cell lines and the tissues from which they came, as well as differentiated cells made from both kinds of stem cells. "If you look with blinders on, they look fairly similar," says Ecker. "But if you zoom in you find different signatures of what an iPS cell is."

The researchers found that rather than being

reset to an embryo-like state, methylation patterns near the tips and centres of chromosomes in the iPS cells resembled those in the adult tissues from which the iPS cells had been derived. This could constrain the types of tissues that the cells are capable of forming. "The reprogramming process, although fascinating, is a fundamentally different way of getting to pluripotency than deriving cells from [embryos]," says George Daley, a stem-cell expert at Children's Hospital Boston in Massachusetts. "We're still looking for reprogramming methods that return cells to the ES-cell-like state," he adds.

The finding that reprogrammed stem cells carry an epigenetic 'memory' dovetails with work published last year by Daley and others comparing mouse iPS and ES cells<sup>2,3</sup>. In mice, however, the methylation differences could be reset, either by continuing to culture the iPS cells or by differentiating the cells again to more specialized cell types. In the human cells, the epigenetic marks lingered even after the iPS cells had been coaxed to form new tissues.

Regardless of their epigenetic differences, neither iPS cells nor ES cells may turn out to be perfect models of tissues in the body. Both cell types seem to harbour genomic abnormalities. In separate work published last month<sup>4</sup>, a team led by Jeanne Loring, a stem-cell researcher at the Scripps Research Institute in La Jolla, found that ES cells tended to contain duplicated chunks of DNA linked to genes associated with self-renewal, whereas iPS cells incorporated extra cancer-causing genes and fewer tumour-suppressor genes. These genomic differences between the two types of stem cells probably result from the culturing techniques used to derive and maintain them.

"When we culture cells outside a normal organism they can acquire features that may not be compatible with life once they go back into an organism," says Richard Young, a stem-cell biologist at the Whitehead Institute in Cambridge, Massachusetts.

The impact of such discrepancies remain unclear, says William Lowry, a stem-cell biologist at the University of California, Los Angeles. "The problem is that we don't know if any of these differences are going to be consequential." ■

1. Lister, R. *et al. Nature* doi:10.1038/nature09798 (2011).

2. Kim, K. *et al. Nature* **467**, 285–290 (2010).

3. Polo, J. M. *et al. Nature Biotechnol.* **28**, 848–855 (2010).

4. Laurent, L. C. *et al. Cell Stem Cell* **8**, 106–118 (2011).



The US National Science Foundation has decided to give the ALMA Vertex Prototype Antenna, which can probe the Universe using sub-millimetre-wavelength radio waves, to an institute in Taiwan.

## ASTRONOMY

# Antenna decision makes waves

*Procedural transparency is at issue as a US agency transfers a high-precision radio dish to an international partner.*

BY EUGENIE SAMUEL REICH

Free to a good home: one 12-metre radio antenna, perfect for high-resolution sub-millimetre-wavelength astronomy. Pick it up yourself; no guarantees. Estimated value: US\$10 million to \$15 million.

It was nearly that straightforward. Last year, the US National Science Foundation (NSF) put out a call for expressions of interest in a prototype antenna that it had funded to test specifications for the Atacama Large Millimeter Array (ALMA), a 66-dish radio observatory now nearing completion in Chile. But what began as an opportunity for some cutting-edge science now has some US bidders crying foul after the NSF told them in early January that it is giving the Alma Vertex Prototype Antenna to an institute in Taiwan.

"We've tried to find out why the NSF made the decision and we've been given only generalities," says Lucy Ziurys, an astrochemist at the University of Arizona in Tucson and the principal investigator for one of the US bids. Any suggestion of improper decision-making

would be sensitive for the NSF at a time when government agencies are bracing for scrutiny from a budget-conscious Congress — and the donation of a major piece of research hardware outside the United States could raise uncomfortable questions.

The antenna is valuable to astronomers because it is designed to probe the Universe using radio waves with wavelengths shorter than 1 millimetre — an under-explored region of the electromagnetic spectrum. Ziurys and her colleagues had proposed to put it on Kitt Peak, 60 kilometres southwest of Tucson, where it would be used to study the composition and dynamics of interstellar clouds, including star and planet formation. Instead, the dish will go to the Academia Sinica Institute of Astronomy and Astrophysics (ASIAA) in Taipei.

Ziurys says that a group convened by the US National Radio Astronomy Observatory (NRAO) — which runs the facility near Socorro, New Mexico, where the antenna is currently located — ranked the University of Arizona's bid above ASIAA's for technical merit. But that was not enough to sway Vernon Pankonin, deputy

division director for astronomical sciences at the NSF, who says that he chose ASIAA as the winning bidder after consulting with an anonymous assessment group internal to the NSF. The group considered the NRAO recommendation, but concluded that ASIAA's bid was superior when several other factors, including scientific merit, were taken into account. The decision was not subject to an external peer review as an NSF grant would be. "It is a transfer of property completely independent of the NSF's grant and award process," says Pankonin.

At the NSF's direction, the original call for bids was issued by the NRAO, whose director, Fred Lo, is chairman of ASIAA's advisory panel and is friends with Paul Ho, director of the institute. Lo acknowledges the friendship and says that he may have discussed the antenna with Ho. He also says that he drafted the call for expressions of interest, which noted that proposals for the dish would be considered not only from the United States, but also from "the communities that form the North American ALMA region (i.e. Canada and Taiwan)". Lo says that the sentence was included at the request of the NSF; Pankonin says the process of deciding to include it was "interactive" between the NSF and NRAO. Lo says that after the call, he stayed out of the decision-making. "Precisely because of the potential charge of conflict of interest, the NRAO was quite careful. It took an objective process," he says.

Christine Boesz, a former inspector-general of the NSF, says that no matter what the final decision was, it could be seen as problematic for a person who could be partial to one bidder to write a call for expressions of interest. "You could raise the question of is it really arms-length decision-making," she says. Zachary Kurz, spokesman for the Republican majority of the House Committee on Science, Space and Technology in the US Congress, says that committee staff are "starting a dialogue with the NSF" about a possible conflict of interest.

Officially, the antenna is being given to the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, which collaborated with ASIAA on the proposal, so on paper it will remain a US asset, says Pankonin. Nevertheless, the proposal makes it clear that ASIAA will take responsibility for the antenna, which it plans to use for very-long-baseline interferometry (VLBI) — a technique in which data from radio telescopes continents apart can be combined to produce high-resolution images. The group's primary targets include the centre of the galaxy M87, which contains the only supermassive black hole beyond the Milky Way whose perimeter could potentially be imaged using VLBI. A location for the antenna has not yet been confirmed, but ASIAA is interested in an NSF site known as Summit Station, at the peak of the Greenland ice sheet, where the cold, dry air would allow the telescope to see even shorter wavelengths than are detectable at Kitt Peak. ■



## POLICY

# China sets 2020 vision for science

*Goals include commercialization of research and emphasis on energy, biomedicine and information technology.*

BY JANE QIU IN BEIJING

China is betting that an ambitious programme of applied research will help to secure its future as an economic superpower. Innovation 2020, unveiled last week by the Chinese Academy of Sciences (CAS), maintains support for basic research. But the plan will place a new emphasis on translating the research into technologies that can power economic growth and address pressing national needs such as clean energy, said Bai Chunli, vice-president of the CAS, at the academy's annual conference in Beijing, where the plan was announced.

Innovation 2020 is an extension of the Knowledge Innovation Programme (KIP) launched by the CAS in 1998. Under the KIP, the academy streamlined its often overstuffed and outdated institutes, attracted outstanding Chinese researchers who had trained abroad, and tightened up the way it evaluated project proposals and performance. But the CAS now needs to support new priorities, says Duan Yibing, a policy researcher at the CAS Institute of Policy and Management in Beijing. China has become a global economic power, and the world's financial crisis has made scientific innovation more important to economic success than ever before, he says. "Things are a lot different now compared to 13 years ago."

Although the budget of Innovation 2020 is yet to be announced, insiders say it will be part of a continuing surge in the nation's science spending (see 'Spend, spend, spend'). Indeed,



CHINA DAILY/REUTERS

China is investing heavily in renewable-energy research as it builds its capacity in, for example, solar power.

the CAS's expenditure on research and development (R&D) in 2009 was about 20 billion renminbi (US\$3 billion), seven times the level in 1998, according to a KIP assessment report also released last week. This year's budget for the National Natural Science Foundation of China will increase by 70%, from 10 billion renminbi last year.

Innovation 2020 will kick off with new projects this year in seven key areas, including nuclear fusion and nuclear-waste management; stem cells and regenerative medicine; and calculating the flux of carbon between land, oceans and atmosphere. Other priority areas include materials science, information technology, public health and the environment.

To coordinate resources better and to foster multidisciplinary research, the academy will set up three research centres for space science, clean coal technologies and geoscience monitoring devices. It also plans to build three science parks — in Beijing, Shanghai and Guangdong province, respectively — to accelerate the conversion of basic research into marketable products, especially in renewable energy, information technology and biomedicine.

Pan Jiaofeng, deputy general secretary of the CAS, says the KIP's track record bodes well for the success of the new programme. By the CAS's reckoning, in 2009, researchers that it funded

published 3.5 times as many papers in journals listed by the Science Citation Index (SCI) as in 1998. Crucially, the number of papers published in the top 1% of SCI journals, as judged by their impact factor, was 12 times that in 1998. The CAS also calculates that research and development by the KIP generated an income of 140 billion renminbi and tax revenue of 22 billion renminbi in 2009 — respectively 19.5 and 14.5 times the levels in 2000.

But the report acknowledges that there is substantial room for improvement. For example, CAS researchers should aim to become leaders of the international scientific community, and shift their focus away from generating as many papers as possible and towards genuine originality and innovation.

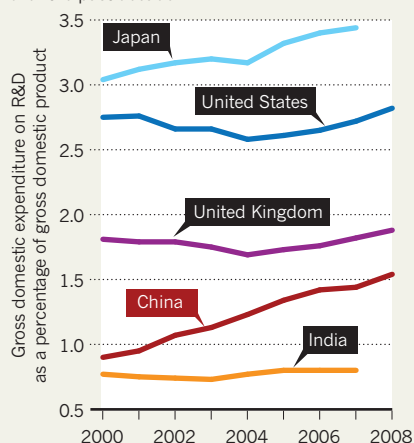
With its emphasis on applied research, the new initiative also "presents a major challenge to the management and organizational capabilities of the academy", says Richard Suttmeier, a science-policy researcher at the University of Oregon in Eugene. He notes that most CAS institutes are focused on academic disciplines and lack the infrastructure needed for commercializing research or directing it towards national needs.

Others think that the emphasis on applied research, national needs and revenue could stifle curiosity-driven research. Without that, says a Shanghai-based researcher who declines to reveal his identity, "it would be very difficult to have genuine innovation". ■

SOURCE: UNESCO

## SPEND, SPEND, SPEND

China's investment in science has risen rapidly over the past decade.



**NATURE.COM**  
For all the latest  
research from China,  
see:  
[www.nature.com/nchina](http://www.nature.com/nchina)





Research commissioner Máire Geoghegan-Quinn faces calls for drastic changes to the EU funding system.

## POLICY

# EU advisers urge funding reform

*The European Commission should free its Framework programme from political interference and red tape.*

BY NATASHA GILBERT

Europe's multi-billion-euro research programme needs significant reform to slash bureaucracy and ensure continued support for cutting-edge science. That's the verdict of the top science advisory group to the European Commission (EC). As the executive body of the European Union (EU), the EC oversees the €50.5-billion (US\$69.3-billion) Framework funding programme.

In a set of unpublished recommendations made to the EC in December, and now seen by *Nature*, the European Research Area Board (ERAB) says that the management of Framework funds should be devolved to "independent institutions at arm's length of Commission and Member States influence". Unless there is a "drastic" change in how the programme operates, it adds, "Europe's ability to compete or cooperate in the global environment will significantly diminish".

The warning comes at a crucial time for the EC, as it prepares to launch a public consultation of stakeholders on the successor to the current Seventh Framework Programme (FP7), Europe's chief research-funding mechanism, which ends in 2013 (see 'Planning a Framework').

Under FP7, the EC organizes research agendas through ten priority themes, such as energy and transport. ERAB suggests that agencies modelled on the European Research Council (ERC) — an EU initiative set up in 2007 to award research grants solely on the basis of excellence — should instead be set up to support these priority research areas.

ERAB says that the EC, as well as the member states of the EU, would continue to have a role in defining the proposed agencies' overall strategy, including research priority areas and their budgets. But the agencies would execute the strategy and determine which proposals would receive

funding, with success judged on the delivery of new discoveries, insights or technologies.

The board acknowledges that some of these research programmes would probably have a higher risk of failure than many FP7 ventures, and says that the agencies would therefore need managers with "considerable responsibilities and powers" who are not restricted by "unnecessary bureaucratic constraints". John Wood, ERAB's chair, declined to comment on the recommendations ahead of their publication.

## HUGE HASSLE

ERAB's recommendations are likely to be welcomed by many of Europe's researchers, who have long deplored the EC's excessive bureaucracy and risk-averse approach to research funding (see *Nature* 463, 999; 2010). "There is a lot of administrative work. Proposals have to be very detailed and precise. This is not always how science works," says Antoine Peters, a molecular biologist at the Friedrich Miescher Institute for Biomedical Research in Basel, Switzerland. Peters adds that he has been put off applying for funding from the programme because "it is such a hassle. I avoid it if I can. I'd rather go for national or local funding."

The European Association of Research and Technology Organisations, a Brussels-based trade group, supports the idea of independent agencies managing research programmes, says Pauline Bastidon, the group's policy officer. In particular, it would like to see an agency, similar to the ERC, in charge of funding for applied research and innovation, she adds.

But Luke Georgiou, vice-president of research and innovation at the University of Manchester, UK, does not think that shifting responsibility to independent agencies is a panacea, and points out that the ERC has still had to battle EC bureaucracy. He proposes retaining the overall shape of the programme but with major simplifications, including more flexibility in calls for proposals.

An EC spokesman declined to comment on ERAB's report, adding that it would be taken into account during the consultation. That process will kick off when the EC releases a green paper outlining its proposals for the next Framework programme on 9 February. But a draft of the green paper, seen by *Nature*, may disappoint research leaders who were expecting to see a set of defined ideas. Instead, it lists 24 broad questions to be addressed in constructing the programme, but offers no firm options. For example, the document asks whether new rules could help to simplify the programme while giving it flexibility, but fails to suggest what these rules would look like.

"The green paper doesn't say anything," says a senior EU science official involved in Framework discussions, who asked not to be named, as commission rules forbid them from commenting on unpublished documents. "It makes me think the commission is not interested in having a proper debate." ■

## PLANNING A FRAMEWORK

Consultation launches on the future of EU research funding	► 9 FEB 2011
Consultation closes	► MAY 2011
Results announced	► JUN 2011
Proposals presented by EC for next Framework programme (FP8)	► END 2011
Proposals negotiated by EC, EU member states and European Parliament	► 2012
FP7 finishes; FP8 begins	► 2013-14

➔ **NATURE.COM**  
For more European science news, see:  
[www.nature.com/regions/europe](http://www.nature.com/regions/europe)

## SPACE SCIENCE

# Swindon, we have a problem

*Britain's space ambitions lack financial fuel.*

BY GEOFF BRUMFIEL

For someone in charge of a brand new space agency, David Williams has had a prosaic year. Rather than standing in front of banks of computers, watching the test launch of a bold new rocket design, the interim director of the UK Space Agency (UKSA) has been shuffling between almost a dozen government entities, arranging the transfer of budget lines and staff members. "Most of what we've done this year has been internal," he says.

The UKSA was announced with much fanfare in March 2010, but its low-key start and lacklustre annual budget of £206 million (US\$332 million; €241 million, see chart) has left some disappointed. "At the moment, we're not happy," says Richard Peckham, the chair of UKspace, the industry's trade association. "But it's still early days."

Britain has a difficult history in space. The country launched its first — and only — entirely home-grown rocket into orbit in 1971, but even before it flew the government had decided to kill the programme to cut costs. More recently, the 2003 failure of the Beagle 2 Mars probe left the nation's pride smarting. Despite these setbacks, the country's private sector has made steady progress. Between 1999 and 2007, the space industry grew by a steady 9% per year and today boasts about £6 billion in annual revenues. Much of that money is made by small satellite manufacturers and telecom companies offering satellite-phone and Earth-observing capabilities.

Until last year, the nation's space agenda — which included environmental satellites and a few collaborative missions with the European

Space Agency (ESA) — was overseen by the British National Space Centre (BNSC), a somewhat toothless organization, also run by Williams. "Different government departments held the budget for the bits of space they were interested in," says Peckham. "The BNSC partnership tried to glue that together."

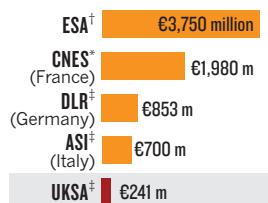
Industry wanted a more powerful agency, along with an increased focus on space as an area of economic growth and innovation. The then Labour government agreed and created the UKSA, along with a space innovation centre at Harwell in Oxfordshire — where a new ESA centre had just been sited — that included a £40-million national space facility.

But as the Swindon-based UKSA prepares to officially open for business on 1 April, it has yet to put forward a clear plan. "We are still waiting for the emergence of a confirmed UK space strategy," says Martin Ditter, head of the ESA centre at Harwell.

The slow start is largely due to the ousting of the Labour government in May 2010. In December, the new coalition government announced that, as part of its austerity programme, the overall UKSA budget for the next four years would be almost flat. That makes it unlikely that the agency will launch its own spacecraft any time soon, although researchers hope that its unifying voice will help Britain to negotiate more prominent roles in ESA science missions.

"Against the backdrop of 25% cuts in government spending, they've probably done a pretty good job," says Peckham. But there is a tinge of disappointment in his voice as he adds, "It doesn't yet look like the empowered agency that we want to see." ■

## SPACE BUDGETS



\*2009 †2010 ‡2011

SOURCE: ASI/CNES/DLR/ESA/UKSA



## LATEST NEWS



Did Viking 'sunstone' help Nordic seafarers to navigate by polarized light? [go.nature.com/pbgqay](http://go.nature.com/pbgqay)

## MORE STORIES

- Arctic Ocean feels the heat rushing in from the Atlantic as sea temperatures rise [go.nature.com/vhcmik](http://go.nature.com/vhcmik)
- Oldest galaxy is lone ranger [go.nature.com/ncvf8f](http://go.nature.com/ncvf8f)
- Insulin-like protein hormone boosts memory [go.nature.com/4i2oyj](http://go.nature.com/4i2oyj)



E. MARCH/CORBIS

Understanding why loneliness can spread through society like a disease is a key question for social scientists.

## RESEARCH POLICY

# Social science lines up its biggest challenges

*'Top ten' crucial questions set research priorities for the field.*

BY JIM GILES

**H**ow can we persuade people to look after their health? Why do moods spread like a contagion? How can humanity increase its collective wisdom?

These are some of the most pressing questions that social scientists should tackle, according to a group of leading scholars in the field who hope that their 'top ten' list will help shape the thinking of researchers and funding bodies for decades to come.

In a parallel effort, the US National Science Foundation (NSF) last week unveiled the results of its own agenda-setting exercise, which asked social scientists to identify "grand challenge questions that are both foundational and transformative".

Both groups say that they ran the exercises because they wanted researchers to step back from immediate research priorities and identify the most significant problems in their field. The results demonstrate the growing ambition of the social sciences to tackle difficult issues in

a quantitative way, addressing problems from equality and wages to wars and health.

The 'top ten' approach was inspired by a list of 23 major unsolved questions compiled by the mathematician David Hilbert in 1900. The Hilbert problems helped to focus the attention of mathematicians throughout the following century. "He laid out the road map for twentieth-century math," says Nick Nash, a vice-president at General Atlantic, an investment firm based in Greenwich, Connecticut. "What if we had a road map for other disciplines?"

In 2008, Nash was studying for an MBA at Harvard University in Cambridge, Massachusetts, when he proposed the road map to Stephen Kosslyn, then the university's dean of social science. Together, they organized a symposium at Harvard last April that gathered together 'big thinkers' to present unsolved questions and to vote on which were the most important. The results are due to be released this week on Harvard's website (see [go.nature.com/wa1hwf](http://go.nature.com/wa1hwf); see also 'Top ten social-science questions'). The site will also include a range of questions

submitted by members of the public.

At the symposium, Emily Oster, an economist at the University of Chicago, Illinois, focused on a perennial challenge for public-health experts: how to get people to adopt healthier behaviours. For instance, persuading people to eat less and exercise more — to control ballooning obesity rates — might be simple in theory; in practice it is extremely difficult.

Because the rewards of behavioural change are often not apparent for years, Oster thinks the answer lies in programmes that offer an immediate pay-off. Preliminary studies have shown, for example, that cash rewards contingent on hitting weight-loss targets can help<sup>1</sup>. Even if payments amount to hundreds of dollars a month and, to prevent a relapse, are continued after dieters have shed their excess pounds, the strategy might save society money by reducing future medical expenses.

The approach is not

**➔ NATURE.COM**  
To suggest more questions for social-science research, comment at: [go.nature.com/oqxqzh](http://go.nature.com/oqxqzh)



1. How can we induce people to look after their health?

2. How do societies create effective and resilient institutions, such as governments?

3. How can humanity increase its collective wisdom?

4. How do we reduce the 'skill gap' between black and white people in America?

5. How can we aggregate information possessed by individuals to make the best decisions?

6. How can we understand the human capacity to create and articulate knowledge?

7. Why do so many female workers still earn less than male workers?

8. How and why does the 'social' become 'biological'?

9. How can we be robust against 'black swans' — rare events that have extreme consequences?

10. Why do social processes, in particular civil violence, either persist over time or suddenly change?

foolproof, though. One recent large-scale experiment, aimed at financially rewarding low-income New York City families for keeping children in school and taking regular medical check-ups, was halted last year after it produced only limited improvements (see [go.nature.com/eunolm](http://go.nature.com/eunolm)). Oster says that researchers need to experiment with different reward systems, as it is not clear how the best system for a particular problem should be chosen.

Nick Bostrom, a philosopher at the University of Oxford, UK, who was also involved in compiling the Harvard list, wants social science to improve society's "ability to get the important things approximately right". He notes that judgements by specialists are often no better than those made by laypeople<sup>2</sup>, and suggests that, rather than relying on individuals, society should develop and exploit new methods for aggregating knowledge.

In financial markets, for example, participants buy and sell shares on the basis of expectations about how the market will move. If enough people play the market, the price of the shares reflects traders' collective beliefs about future events. It is also possible to create artificial markets in which traders buy and sell shares related to specific events, such as a politician being elected. These markets also reflect traders' beliefs about outcomes, and can be good forecasting tools<sup>3</sup>. Bostrom would like to see such 'prediction markets' trialled more widely; they could assist with corporate decisions, such as whether to replace a company chief executive, he suggests.

Harvard social scientist Nicholas Christakis

hopes to understand how physiological and psychological attributes, such as obesity and loneliness, can spread through a social network like a contagious disease, a phenomenon he has studied with James Fowler at the University of California, San Diego<sup>4</sup>. If one of your friends becomes obese, for example, your own chances of putting on weight increase. Christakis notes that there is unlikely to be a single theory that links social and biological factors. In the case of obesity, it may be that having an overweight friend somehow normalizes the idea of gaining weight. And preliminary work in poor neighbourhoods in Chicago suggests that loneliness and fear of crime can alter levels of stress hormones, which in turn can affect people's risk of cancer.

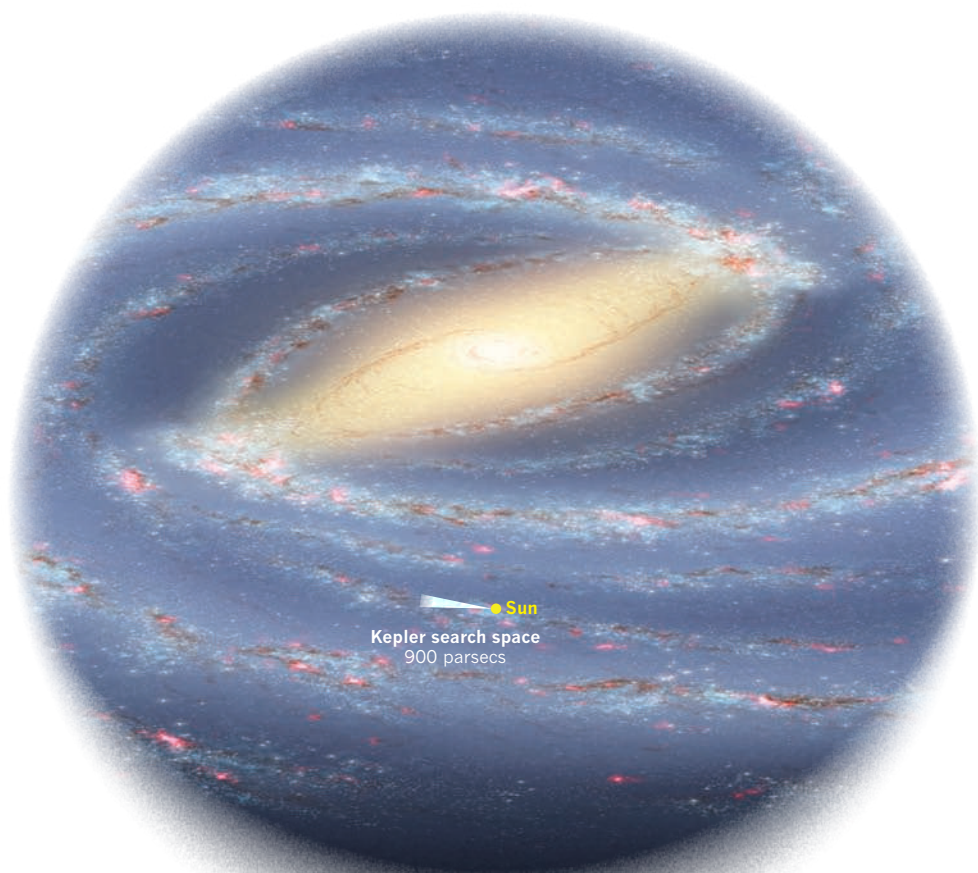
Nash and Kosslyn hope that drawing attention to difficult and important problems will motivate young researchers to work on them, just as young mathematicians were attracted to the Hilbert problems. "Nothing would make us happier than to see

future grant applications that mention the 'Harvard problems,'" says Nash. The similarity with the NSF's own exercise is not a bad thing, adds Myron Gutmann of the foundation's directorate for social, behavioural and economic sciences in Arlington, Virginia. "I'm delighted by [the Harvard exercise]," he says. "It allows us to look for repeated themes."

The NSF received more than 240 responses to its request for forward-looking ideas, which Gutmann plans to discuss with his advisory committee. "I can imagine that in the next two years we will identify a few ideas that seem especially important and invest in them, in the form of pilot projects and planning grants, with an idea that these investments will position us to make more significant investments 5–10 years from now," he says.

Cary Cooper, chair of the Academy of Social Sciences in London and a psychologist at Lancaster University, UK, is enthusiastic about the Harvard list. If funding were available to support work on the problems, he says, young researchers might feel confident enough to eschew simpler questions. Cooper adds that he will consider asking Britain's Economic and Social Research Council to run a similar exercise. ■

1. Volpp, K. G. et al. *J. Am. Med. Assoc.* **300**, 2631–2637 (2008).
2. Tetlock, P. E. *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton Univ. Press, 2006).
3. Berg, J. E., Nelson, F. D. & Rietz, T. A. *Int. J. Forecasting* **24**, 285–300 (2008).
4. Christakis, N. A. & Fowler, J. H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives* (Little, Brown, 2009).



# BEYOND THE STARS

*Launched in 2009 to seek out worlds beyond the Solar System, the Kepler mission is exceeding expectations. Is it closing in on another Earth?*

BY EUGENIE SAMUEL REICH



itting for an interview in his office at the Harvard-Smithsonian Center for Astrophysics (CFA) in Cambridge, Massachusetts, the normally voluble astronomer Dimitar Sasselov looks nervous. Asked for his favourite among the many potential planets discovered by NASA's Kepler planet-finding mission, for which he is a co-investigator, he hesitates, then sidesteps the question entirely. "Personally, I'm already beyond that point. It's not one. It's not a single planet. It's a whole family."

Sasselov has good reason to be wary: his public lecture last July at the Technology, Entertainment and Design 2010 conference in Oxford, UK, earned him a stern rebuke from his colleagues on the mission. Not only had he presented numbers for possible planets greater than those released officially by the team, they said, but he had also used a careless phrasing that resulted in a raft of headlines proclaiming — incorrectly — the discovery of hundreds of other Earths.

That hullabaloo became a distant memory this week, as the mission released 400 candidate systems, adding to the 306 released last June. Along with the candidates came a bunch of confirmed planets. The latest finds, posted on NASA's website last month

**The Kepler space telescope is exploring a sliver of the Milky Way some 900 parsecs (about 3,000 light years) deep.**

NASA/JPL-CALTECH/R. HURT; SSC-CALTECH

(see [go.nature.com/aejd15](http://go.nature.com/aejd15)) and published in *Nature* this week<sup>1</sup>, include a rocky planet orbiting so closely to its star that its starlit side must be a seething sea of lava; and a planetary system containing several large, rocky or icy planets in orbits of tens of days, just one order of magnitude faster than Earth's 365-day cycle. "It's very exciting. It's a type of system we haven't seen before," says Jack Lissauer, a space scientist and Kepler co-investigator based at the NASA Ames Research Center in Moffett Field, California, and a lead author on the paper in *Nature*.

Nonetheless, most of the Kepler scientists continue to be cautious. By watching the light from some 150,000 stars for the dimming that could signal a planet crossing in front of them, Kepler is extraordinarily efficient at finding possible planets. But Kepler has yet to find another Earth — a small, rocky planet with an orbit of a few hundred days and well inside the habitable zone in which water can exist and life can arise. That is for a fundamental reason; the blips that Kepler detects show only the radius, and not the mass, of an observed planet, which means that the density and composition generally remain unknown.

Moreover, the scientific objective of the Kepler mission is not to discover Earth-like planets. Instead, it is to estimate the fraction of Sun-like stars that have Earth-like planets — statistics that could greatly enhance astronomers' understanding of how planetary systems form. Determining which of the blips correspond to planets — rather than systems of stars in which one is eclipsed, causing a similar dimming — is what the researchers spend most of their time on, says William Borucki, a space scientist and Kepler principal investigator at NASA Ames. The only way to do that, he says, is the hard way: painstakingly sorting the real signals from the false positives.

## OBSERVATIONAL BIAS

Until Kepler, the leading detection method used to discover exoplanets — planets outside the Solar System — was much more likely to find giant planets, resulting in a sampling bias. Known as radial velocity or Doppler spectroscopy, the method depends on identifying the shift in a star's spectral lines as it wobbles around a mutual centre of gravity with a planet. The larger the planet and the closer to the star it lies, the faster the star's movement towards and away from Earth, and the easier it is to detect the shift in the spectral lines. Almost all of the planets found by this technique have been larger than Jupiter and very close to their stars, sometimes completing an orbit in just a few days.

An alternative method was presented in 2000, when CFA astronomer David Charbonneau and his colleagues, working from a shed in a car park outside the National Center for Atmospheric Research in Boulder, Colorado, observed a planet passing across — or transiting — the face of its parent star<sup>2</sup>. Within days, another group had made a similar observation<sup>3</sup>. In this case, the researchers were confirming a transit predicted for a planet, HD 209458b, that had been spotted using the radial-velocity method.

Before long, planets were being detected by their transits alone. Those early detections also yielded large, close-in planets, which were easier to see because they obscured larger portions of their host stars than their Earth-size counterparts would do. But researchers were thrilled to realize

that, in principle, a space telescope could be made sensitive enough to see the transits of Earth-sized planets in Earth-like orbits — and the idea of Kepler was born.

Kepler was designed as a 0.95-metre-diameter space telescope that would detect exoplanets by monitoring variations in the light from stars. Unlike most ground-based telescopes and the French Space Agency's COROT planet-finder mission, which monitor targets for months at a time at most, Kepler was intended to stare at the same, fixed field of view for 3–4 years. This field encompasses 150,000 stars in the Cygnus and Lyra constellations, chosen because of a prevalence of Sun-like stars. The commitment to the same star field made Kepler

unique in being able to capture three or four repeat observations of transits by small planets in yearly, Earth-like orbits, even if it wouldn't be able to determine their mass and composition.

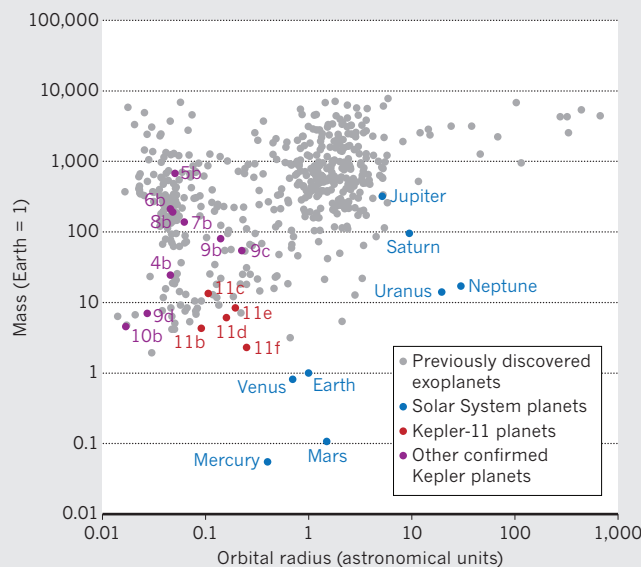
The spacecraft was launched in March 2009, and the Kepler scientists announced their first few planets the following January. "We were just skimming the cream off the top," says Natalie Batalha, an astronomer at San Jose State University in California and Kepler's deputy science team leader. At that point, the short timescale over which the mission had been operating continued to favour the discovery of giant planets with fast orbits that were too close to their host stars to be habitable. These included five giant planets with orbits of between 3.2 days and 4.9 days<sup>4</sup>. But the 306 planetary candidates released a few months later told a different story. Most of these correspond to planets that are Neptune-sized or smaller, and nearly 40 are smaller than twice Earth's size<sup>5</sup>. If confirmed, Batalha estimates, five would correspond to planets orbiting within the habitable zones of their stars.

The process of turning a candidate into a confirmed planet is tortuous. Every month, pixels representing a continuous flux of light captured from the target stars are downloaded from the spacecraft to computers at NASA Ames, where they are converted into light curves — graphs showing the intensity of light from the star as it changes with time. The software flags up about 2,000–3,000 dips in light curves automatically, and these are then sent to a committee headed by Batalha. Those not rejected as obvious false positives — owing to instrument noise, for example — are assigned a Kepler object of interest (KOI) number. Mission scientists estimate that 50% of the KOIs are real planets, but they have been able to confirm only 15 of the 306 KOIs announced so far (see 'No place like home') — including those published in this issue.

The most obvious way to rule out a false positive is to detect the planet using another method. Charbonneau, now a participating scientist on Kepler, is working to follow up the KOIs with NASA's Spitzer Space Telescope, which is sensitive to infrared radiation. It is ideal for ruling out a type of false positive known as a blend, which consists of a much brighter star in the same line of sight as two dimmer, orbiting stars, so that one occasionally eclipses

## NO PLACE LIKE HOME

Kepler looks for exoplanets with masses and orbits similar to Earth's. Most of the discovered exoplanets are very different, but the Kepler-11 planets (estimates for only five are exact enough to plot) are among the closest yet.





the other. To Kepler, a blend can look like the transit of a Jupiter-sized planet, says Charbonneau — but not to Spitzer, because the three stars will have different proportions of their light in the infrared and visible wavelengths.

Kepler data alone can confirm a candidate when it is part of a system of several planets. Last year, for example, researchers found a planetary system with a pair of planets transiting the same star, at almost regular intervals. “We started to lavish more individual attention on the system once we saw the transits were varying,” says Matthew Holman, an astrophysicist at the CFA and a member of the Kepler team. It was a sign that the planets were real: planets can vary by as much as several minutes per orbit if they are interacting gravitationally with one another. In a rapid orbit, this is the equivalent of Earth’s revolution around the Sun changing by a few hours each year. The new planets, dubbed Kepler-9b and -9c, had radii about 0.8 times the size of Jupiter’s, and orbits of 19 days and 39 days, respectively<sup>6</sup>. By modelling the gravitational interactions of the planets, the team calculated that their probable masses are similar to that of Saturn.

Knowing the mass and radius of each planet helped the astronomers to estimate that the composition of the planets was hydrogen- and helium-rich, making them very much like the gas giants Saturn and Jupiter.

Kepler-9 was the first system in which several planets were found to transit the same star. The paper on page 53 (ref. 1) announces a second, Kepler-11, in which as many as six planets transit, with orbital periods of 10, 13, 22, 31, 46 and 118 days, and masses between 2.3 and more than 300 times that of Earth. Although the outer four are gas giants, the inner two could be ice giants like Neptune. Or they could be super-Earths — planets several times larger than Earth but consisting of a mixture of rock and gas. The team was surprised to see as many as six planets transiting in the same system, says Lissauer, and further astounded to find the inner planets so densely packed that, were they any more so, their orbits would not be stable. “It’s an amazing system,” he says.

## SUBTLE EFFECTS

Kepler-11b–g come hot on the heels of the January announcement of Kepler-10b, a dense planet circling a Sun-like star in a 0.84-day orbit. In this case, there were no observable transit-timing variations, but because of the closeness to the star, the researchers were able to confirm the planet using ground-based radial-velocity observations that also showed the planet’s mass. At 4.6 times Earth’s mass and only 1.4 times its radius, the planet is dense enough to be unambiguously rocky — although its closeness to the star means that one side will be constantly molten rock. After careful studies of the host star and analyses of a year’s worth of data, the Kepler team was able to pick out not only the dimming due to the transit, but also the cycle of brightening and dimming as the orbiting planet alternately showed its day and night sides towards Earth. “It’s phenomenal” that such a subtle effect was detectable, says Batalha. It also shows what unexpected insights can be gleaned from Kepler’s observations. “This is our first data point down in

the rocky regime,” says Batalha, “it’s a huge milestone.”

The Kepler mission has detected other planets for which the mass cannot be determined. One example is Kepler-9d, a super-Earth found in the same system as Kepler-9b and -9c. Computer simulations by Guillermo Torres, an astronomer at the CFA, and his colleagues showed that eclipsing stars could not produce as good a fit to the observed light curve as the match produced by a super-Earth planet passing in front of the Kepler-9 star. That, says the team in a paper in *The Astrophysical Journal*<sup>7</sup>, is the first validation of a planet using a general method that could be applied to any of the Kepler candidates, even those that don’t show transit-timing variations and are too far from the star to be studied using the radial-velocity method. “The probability of a planet is higher than the probability of a false positive,” says Torres, “there is a statistical argument.” The method was used again to validate the sixth of the Kepler-11 candidates, Kepler-11g, a gas giant with an orbit of 118 days, far enough from the rest of the cohort that any transit-timing variations are too subtle to have been observed.

## THE KEPLER LEGACY

Because of the time needed for repeat observations of planets in Earth-length orbits, it will be years before the Kepler researchers

can establish the frequency of planets in the cosmos. But that hasn’t prevented other scientists from making preliminary estimates. In 2010, for example, a group led by Andrew Howard, an astronomer at the University of California, Berkeley, took the size distribution of planets found by the radial-velocity method and, by extrapolating to lower masses, predicted that Kepler will find that roughly 22% of stars are orbited by Earth-size planets<sup>8</sup>.

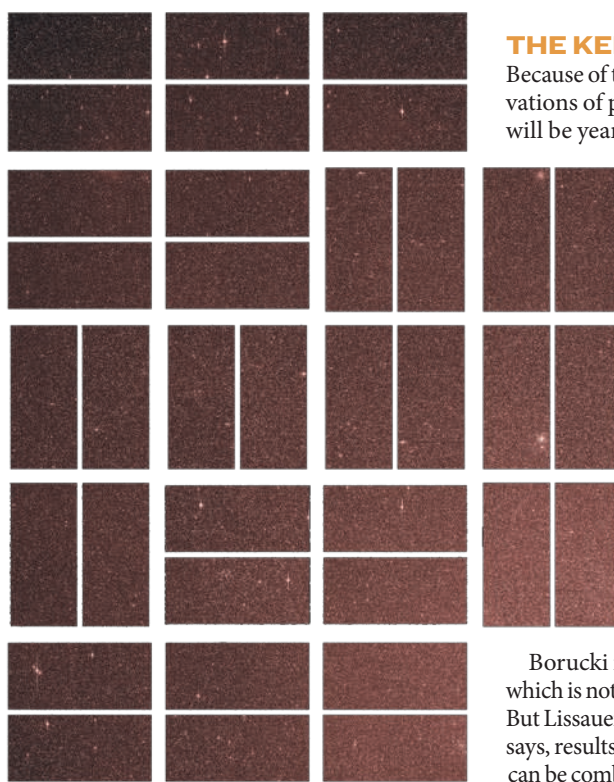
Borucki is sceptical. “They extrapolated, which is not a mortal sin but it’s close,” he says. But Lissauer is more sanguine. Ultimately, he says, results from the radial-velocity method can be combined with results from the transit method to produce a measured frequency of planets at different sizes, masses and compositions — from rocky Earths to gaseous Jupi-

ters. And those data, in turn, will be invaluable for helping astronomers to understand the origin and evolution of planetary systems throughout our Galaxy. “There’s a whole load of good science in there,” says Lissauer.

The promise, says Batalha, is that Kepler will deliver a massive roster of objects for future generations to follow up on. “Kepler will leave this legacy. People are going to use these data for decades,” he says. ■ **SEE EDITORIAL P.5**

**Eugenie Samuel Reich** is a reporter for Nature based in Boston.

1. Lissauer, J. L. *et al. Nature* **470**, 53–58 (2011).
2. Charbonneau, D., Brown, T. M., Latham, D. W. & Mayor, M. *Astrophys. J.* **529**, L45–L48 (2000).
3. Henry, G. W., Marcy, G. W., Butler, R. P. & Vogt, S. S. *Astrophys. J.* **529**, L41–L44 (2000).
4. Borucki, W. J. *et al. Science* **327**, 977–980 (2010).
5. Borucki, W. J. *et al. Preprint at* <<http://xxx.lanl.gov/abs/1006.2799>> (2010).
6. Holman, M. J. *et al. Science* **330**, 51–54 (2010).
7. Torres, G. *et al. Astrophys. J.* **727**, 24 (2011).
8. Howard, A. W. *et al. Science* **330**, 653–655 (2010).



Kepler’s full field of view, which encompasses 100 square degrees of sky and 150,000 stars.

**NATURE.COM**  
For more on  
exoplanets, visit:  
[go.nature.com/ibz7ft](http://go.nature.com/ibz7ft)



# EXOPLANETS ON THE CHEAP

*The search for planets outside our Solar System will always be pricey. But creative solutions are proving that it no longer has to break the bank.*

BY LEE BILLINGS



Astronomers searching for planets around stars other than the Sun have had much to celebrate over the past decade. The number of confirmed ‘exoplanets’ has soared from about 50 to more than 500 in that time. And although none of these planets closely resembles Earth, NASA’s Kepler space telescope, launched in 2009, is now delivering candidates from distant stars by the hundreds — some of which may prove to be very Earth-like indeed (see page 24).

The exoplanet search itself has been wildly successful, but not so the searchers’ quest for multibillion-dollar follow-up missions. Hopes for ambitious spacecraft such as a Space Interferometry Mission or Terrestrial Planet Finder have been dashed as missions have been cancelled or postponed owing to a combination of sluggish economic growth, deep cuts to space-science funding and programme difficulties with NASA’s James Webb Space Telescope (JWST).

In response, the planet-hunting community has got creative, devising ways to maximize the science and minimize the costs. An Exoplanet Task Force jointly commissioned by NASA and the US National Science Foundation accordingly issued a report<sup>1</sup> in 2008, supporting a new strategy for exoplanet research. Rather than waiting for the launch of costly, dedicated planet-hunting spacecraft, it calls for astronomers to press ahead with cheaper, ground-based surveys to discover worlds orbiting nearby stars, which appear brighter to us than do those farther away, and so are easier to study. The hope is that such low-cost surveys will yield at least a few worlds that can be studied using space-based resources such as the JWST. Such facilities would allow astronomers to spectroscopically search the exoplanets’ atmospheres for ingredients such as carbon dioxide, water vapour and perhaps methane, oxygen and other trace gases, which could indicate that life is present. A 2010 report from the European Space Agency reached nearly identical conclusions<sup>2</sup>.

“The planets are out there, and it’s relatively inexpensive to go after them,” says Greg Laughlin, an astrophysicist at the University of California, Santa Cruz, who served on the NASA/National Science Foundation task force. “There’s an economic inevitability to this.”

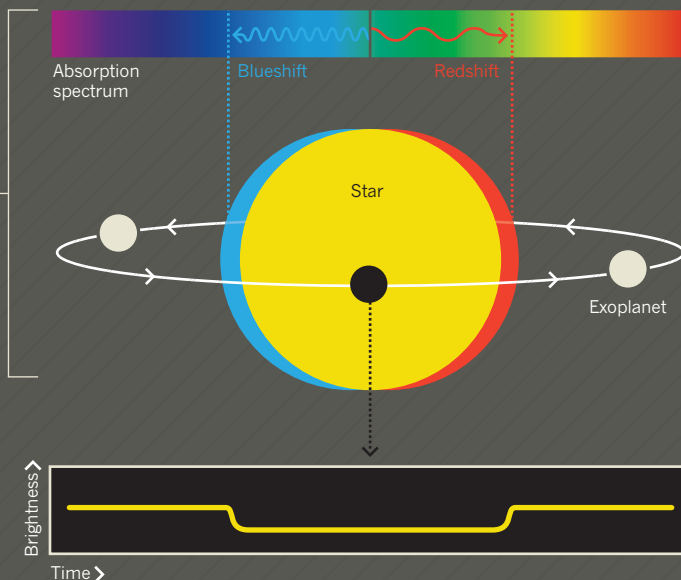
Of the many ideas that astronomers have come up with for conducting exoplanet searches on the cheap, five stand out. ▶

# PLANET-HUNTING FOR BEGINNERS

Of the handful of techniques for finding exoplanets — planets orbiting stars other than our Sun — two are by far the most productive.

## RADIAL VELOCITY

Radial velocity is the motion of a star caused by the gravitational influence of its orbiting planets. It can be measured through increases (blueshifts) or decreases (redshifts) in the frequency of light the star emits. Radial-velocity measurements can detect only planets whose orbits tug the star towards and away from the observer. The exact orbit of an exoplanet is hard to determine, so radial-velocity measurements let researchers deduce only the time a planet takes to orbit the star (its orbital period), how its orbit deviates from circular and its minimum mass. Radial velocity is most sensitive to massive planets with short orbital periods.



## TRANSITS

When a planet crosses, or 'transits', the face of its star, it dims the star's light by a small but detectable amount. The probability that any planet's transit will be visible from Earth is low, and is dictated by the ratio of the diameter of the star to the diameter of the planet's orbit. Large planets with short-period orbits of small stars are most likely to be seen transiting, and lots of stars must be surveyed for any transits to be found. Transits let researchers deduce the radius of a planet and its orbital period. Astronomers can sometimes study a planet's atmosphere as starlight filters through or reflects off it. This gives information on atmospheric composition, temperature and cloud formation.

## ► M-DWARF TRANSIT SURVEYS (US\$2 MILLION)

Central to the strategy is a focus on cool, red 'M-dwarf' stars close to the Solar System. Not only are there lots of them — M-dwarfs are the most abundant kind of star in the Milky Way — but they are much smaller and dimmer than the Sun, having less than half its mass. So any M-dwarf planet passing in front of the star, or 'transiting' it, would block a larger fraction of the light than it would of a larger star and would be easier to detect (see 'Planet-hunting for beginners'). The comparative size of the transiting planet's silhouette would also make it easier for telescopes to gather light filtering through its atmosphere for spectroscopic analysis.

The first, and so far most successful, search for potentially habitable planets transiting M-dwarfs is the MEarth Project (pronounced 'mirth'): a cluster of eight 0.4-metre robotic telescopes at the Whipple Observatory on Mount Hopkins in Arizona. Unlike all previous transit surveys, which stare at a fixed patch of sky rich with stars, MEarth targets 2,000 nearby M-dwarfs; only if one of these displays a candidate transit will all telescopes observe it at once. The project is headed by David Charbonneau, an astronomer at Harvard University in Cambridge, Massachusetts, and was designed mainly by Philip Nutzman, now a postdoctoral researcher in astronomy at the University of California, Santa Cruz. MEarth announced the discovery of its first transiting planet in 2009 — a world dubbed GJ 1214 b, after the M-dwarf star it orbits some 13 parsecs from Earth<sup>3</sup>. The planet is too large and hot to harbour life as we know it, but was found in only the first six months of MEarth's proposed three-year running time, and so far remains the most easily studied Earth-like exoplanet known. A spectroscopic study of GJ 1214 b, undertaken last year at the European Southern Observatory (ESO) in La Silla, Chile, showed that the planet's upper atmosphere is either very hazy or is composed of water vapour<sup>4</sup>.

"MEarth shows that for a relatively modest investment of US\$1 million or \$2 million, you can put together a ground-based survey capable of finding habitable-zone super-Earths," says Charbonneau, referring to rocky planets that are larger than Earth and orbit their stars at a distance at which water can exist as a liquid. "The answer to the question being asked is certainly worth a lot more than that."

By October 2011, Charbonneau and his colleagues hope to have a copy of

MEarth operating in Chile, where it will see parts of the sky not visible from Arizona. Several other M-dwarf transit searches are also under way, notably at the 0.6-metre TRAPPIST telescope in La Silla, and at the 1.22-metre Samuel Oschin Telescope at Palomar Observatory in California.

## NEAR-INFRARED SPECTROMETERS (US\$5 MILLION)

Although transit observations are a good way to determine the radii and orbital periods of exoplanets, other techniques, such as spectroscopy, are essential for learning more.

Particularly important is the radial-velocity technique, the planet-hunting method that has produced the most hits so far. An orbiting planet tugs its star to and fro, generating periodic shifts in the wavelengths of the light the star emits; measurements of these shifts can not only allow independent confirmation of an exoplanet's existence, but also provide estimates of its mass.

The method presents another good reason to focus on M-dwarfs. Earth's motion around the Sun causes the star to wobble with a radial velocity of some 10 centimetres per second over the course of a year — a tough signal for any alien astronomers to detect. But if a planet the size of Earth were located in the habitable zone of an M-dwarf, much closer to the star, its radial-velocity signature would be a metre per second, much easier to see. Unfortunately, M-dwarfs shine most brightly with infrared and near-infrared light, so that is the region of the electromagnetic spectrum in which planet-hunters must search — but astronomers have yet to build the infrared spectrometers required for such precise measurements. So only a handful of the myriad M-dwarfs close to the Solar System have been surveyed for habitable planets.

Worse, planet-hunting near-infrared spectrometers are more costly than their optical counterparts, owing to basic physics: infrared photons don't have enough energy to easily excite electrons in an off-the-shelf

silicon detector. So the instruments rely instead on detectors made from expensive, exotic materials such as indium gallium arsenide or mercury cadmium telluride, and must be either cryogenically cooled or thermally insulated against background

► **NATURE.COM**  
For more on  
exoplanets, visit:  
[go.nature.com/ibz7ft](http://go.nature.com/ibz7ft)



infrared radiation. The most cutting-edge near-infrared spectrometer is ESO's Cryogenic High-resolution Infrared Echelle Spectrograph (CRIRES), which cost some €10 million (US\$13.6 million) to build.

Yet prices are dropping, and several major spectrometers may debut in the next few years, if they receive sufficient funding. Among them are the Calar Alto High-resolution Search for M-dwarfs with Exo-Earths with a Near-infrared Echelle Spectrograph (CARMENES), a German-Spanish instrument slated for the Calar Alto observatory in Spain; a near-infrared spectropolarimeter (SPIRou) on the Canada-France-Hawaii Telescope in Hawaii; and a US project, the Habitable Zone Planet Finder, slated for the Hobby-Eberly Telescope at the McDonald Observatory in Texas.

## LASER FREQUENCY COMBS (US\$100,000)

Another hurdle to the progress of M-dwarf planet discoveries is more subtle: the need for better ways to calibrate the spectrometers. The minuscule spectral-line shifts caused by an orbiting habitable planet can all too easily be mimicked by fluctuations in the stability of the instruments themselves. The obvious solution is to generate a reference spectrum with which the observations can always be compared. But the spectra generally used to calibrate optical radial-velocity surveys — those from iodine or a thorium-argon mix — don't produce usable calibration lines in infrared.

A number of other elements and mixtures are under investigation for calibration. But planet-hunters are most excited about an ultra-high-precision technology known as the laser frequency comb. At the core of such a device is a laser that emits rapid pulses that can be tuned across a wide range of wavelengths. Plotting the frequency of such a pulse train gives a distinct series of regular-wavelength peaks that resemble the teeth of a comb. When those pulses are fed through a spectrometer and synchronized with the ticking of an atomic clock, it becomes a powerful calibration source for spectroscopic measurements.

Efforts are under way to test laser combs on spectrometers at observatories. In late 2009 and early 2010, for example, a laser comb developed at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, was linked to an optical spectrometer at the Whipple Observatory and used to calibrate spectroscopic measurements of a known planet-hosting binary star, HD 189733. In mid-2010, at the Hobby-Eberly Telescope, a comb developed at the US National Institute of Standards and Technology and mounted on a near-infrared spectrometer from the University of Pennsylvania, Philadelphia, obtained radial-velocity measurements of the planet-hosting star Upsilon Andromedae. And in December 2010, at ESO, a comb from the Max Planck Institute of Quantum Optics in Garching, Germany, made radial-velocity measurements of an exoplanet that were, for the first time, more accurate than the previous champion, thorium-argon calibration. The results of these tests are unpublished. If all goes well, combs from each team will grace next-generation spectrometers at major observatories in this decade.

## RADIAL-VELOCITY OBSERVATORIES (US\$50 MILLION)

Even armed with laser combs, planet-hunters could still be undone by the stars themselves, whose surface motions can masquerade as radial-velocity signals. "A star reverberates like a bell, with millions of modes of harmonic oscillations covering its surface, a bit like the weird patterns you get from putting sand on a vibrating drum head," says Steven Vogt, an astronomer at the University of California, Santa Cruz. "A few of these modes don't average out across the surface of the star, and they give you oscillations that can show up as noise in your observations."

The technique that has emerged to counter such noise sources is to average together 10–15-minute time-exposures of the star taken on consecutive nights over a period of weeks. Stéphane Udry, an astronomer at the University of Geneva in Switzerland, who hunts for planets at La Silla, says that it works. "We have a little sample of ten nearby stars we've begun

following in this way, and we've already found planets around three of them," he says. "But a lot of observations are needed, because the stars are unlikely to have only one planet. So we have to cover all the potential periods for multiple planets, which takes time. As you try to make your measurements more precise, it quickly becomes expensive."

So expensive, in fact, that Vogt says the best way to reduce the long-term cost is to spend more money in the short-term on building radial-velocity-dedicated observatories. "The coin of the realm is observing nights," he says. "It's not new technology; it's not laser combs or some newfangled near-infrared spectrometers that can take advantage of M-dwarfs. Take \$50 million, which is chump change in the NASA regime, build a 6–8-metre telescope with enough light-gathering power to reach a large fraction of the nearest M-dwarfs, put a nice spectrometer on it and dedicate it to this work every single night of the year. You'd have these planets pouring out of the sky."

Vogt and his colleagues have built a demonstration project, the Automated Planet Finder (APF): a 2.4-metre robotic telescope paired with a high-efficiency spectrometer at Lick Observatory on Mount Hamilton, California. The APF, according to Vogt, is "built and bred only to find short-period rocky planets" around nearby stars, including the brightest M-dwarfs in the sky. The project is now in its final installation phase, with commissioning scheduled for this month. Vogt expects it to rapidly discover a bevy of small, rocky worlds.

## EXOPLANETSATS (US\$250,000 EACH)

Radial velocity's most important role in the future may be helping to verify and study promising transit discoveries. "Transit searches are the most advantageous technique giving access to terrestrial planets in the habitable zones of stars," says Udry. "We cannot beat that."

Astronomers are already brainstorming successors to the Kepler mission, which would carry out transit surveys of nearby stars looking for worlds with the potential for life. In the meantime, a much cheaper proposal is the ExoplanetSat programme being developed by Sara Seager, an astronomer at Massachusetts Institute of Technology in Cambridge, and her team. The idea is to build on the existing framework for 'CubeSats' — miniaturized satellites, made up of varying numbers of cubes 10 centimetres on each side, designed to hitch low-cost rides into orbit on rockets launching larger spacecraft.

Seager's plan calls for a fleet of dozens of CubeSats, each targeting an individual star and containing a small telescope and guidance equipment. Outside Earth's atmosphere, which interferes with observations, such a payload could detect transiting Earth-sized planets in the habitable zones of nearby Sun-like stars.

Seager admits that engineering such 'nano-satellites' to have the necessary stability and thermal control will be challenging. But she and her team hope to launch a functional prototype as early as 2012, with subsequent satellites launching for as little as \$250,000 apiece — a bargain-basement price for a space-science mission. "On one hand, this seems risky because the probability of finding a transiting Earth-sized planet in the habitable zone of a nearby star is currently estimated at 1 in 200," she says. "On the other, these satellites are modular and relatively cheap; launching one has low risks and what may be very high returns. Basically, this could be MEarth in the sky." ■ [SEE EDITORIAL P.5](#)

**Lee Billings** is a freelance writer based in New York.

1. The Exoplanet Task Force *Worlds Beyond: A Strategy for the Detection and Characterization of Exoplanets* (NASA Astronomy and Astrophysics Advisory Committee, 2008).
2. Exoplanet Roadmap Advisory Team *A European Roadmap for Exoplanets* (ESA, 2010).
3. Charbonneau, D. *et al. Nature* **462**, 891–894 (2009).
4. Bean, J. L., Kempton, E. M.-R. & Homeier, D. *et al. Nature* **468**, 669–672 (2010).

**"THE PLANETS ARE  
OUT THERE, AND IT'S  
RELATIVELY INEXPENSIVE  
TO GO AFTER THEM."**

# COMMENT

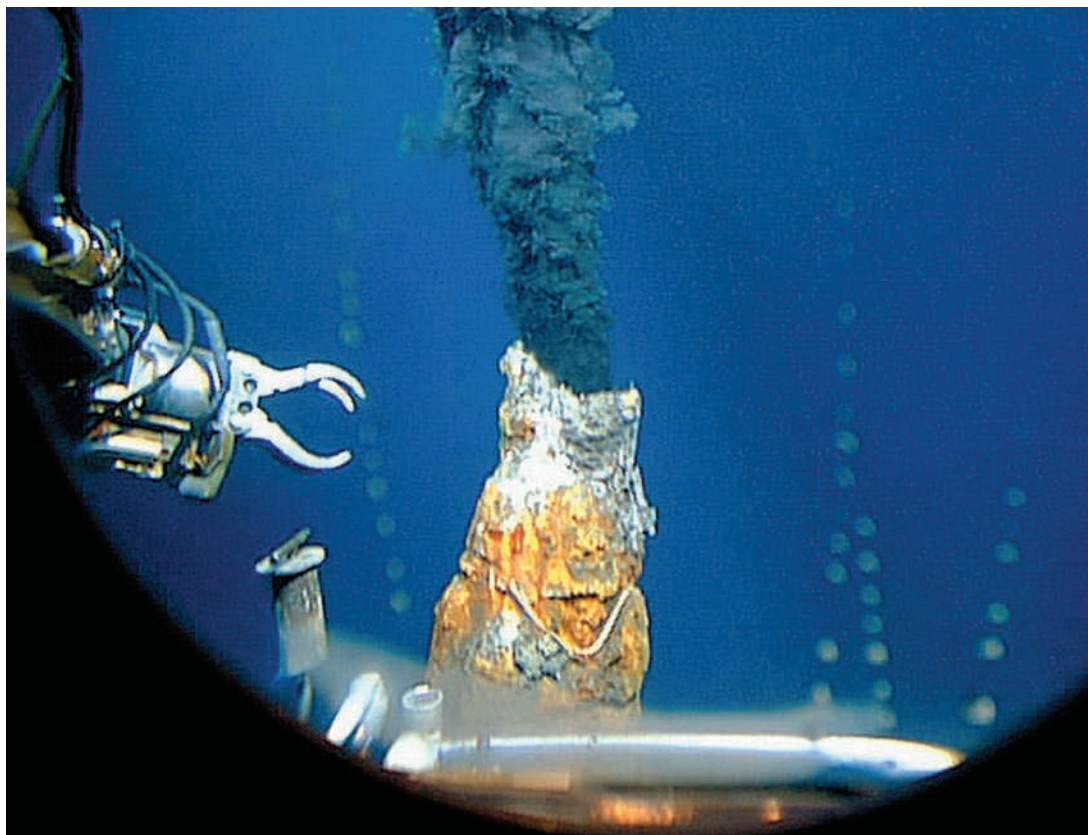
**ASTRONOMY** Passionate account of the demotion of Pluto **p.34**



**BOTANY** Plant-hunting and cross-dressing on the high seas **p.36**

**PUBLISHING** Invisibility of negative results skews the scientific record **p.39**

**OBITUARY** Eugene Goldwasser, who hunted down EPO hormone, remembered **p.40**



Deep-sea vents are underwater hot springs, home to unique life forms and metal-rich minerals.

## Tighten regulations on deep-sea mining

Extracting minerals from sea-floor vents should not go ahead without a coherent conservation framework, argues **Cindy Lee Van Dover**.

**D**eposits of gold ore found along the Salmon River in the northwestern United States during the 1860s attracted explorers to the hot mineral springs of the Yellowstone Basin. Soon after, speculators moved in intending to fence and claim the land containing the hot springs. Instead, by 1872, the Yellowstone geyser basin was set aside as the world's first national park. Remarkably, policy-makers in Washington DC, whose only knowledge of Yellowstone was based on photographs, paintings and stories, swiftly saw fit to leave this wilderness pristine for future generations.

In the late 1970s, geologists discovered analogous, mineral-rich hot

springs in volcanically active areas of the floor of the Pacific Ocean<sup>1</sup> (see map). These deep-sea hydrothermal vents support bacteria that use chemicals in the vent fluids to generate cellular energy. The bacteria feed luxuriant communities of beautiful and strange invertebrates in an otherwise barren seascape. Scientists studying vents have gained insights into the cooling of Earth's interior, ocean chemistry and the extremes at which life can exist on Earth and potentially elsewhere in the Universe. Some national governments, such as those of Canada, Portugal, Mexico and the United States, have introduced marine parks to protect vent fields of particular scientific interest within 200 nautical miles of their ►

P. HICKEY/WHOI



► coastlines. But most vents are found in international waters, where there is little environmental oversight of deep-sea habitats, or in the territorial seas of countries with nascent or non-existent conservation policies that apply to deep-sea hydrothermal vents.

With commodity prices on the rise, mining of mineral deposits at deep-sea vents looks set to begin in the next few years. As a scientist who has studied hydrothermal vents almost since their discovery, and as one passionate about the exquisite organisms that thrive there, I would prefer that sea-floor hot springs remain pristine — deep-sea Yellowstones — untouched by mining. But I recognize that the scientific values I assign to hydrothermal vents must be weighed against other values, including economic ones.

With mining likely to be inevitable, scientists need to promote conservation at every level — from international governance agencies to individual mining companies. To that end I am working with Nautilus Minerals, a deep-sea mining company headquartered in Toronto, Canada, undertaking research that informs its environmental management strategies. In return I am able to tackle research questions that would otherwise be out of reach owing to high costs of field sampling in the deep sea. Some may consider such an alliance a Faustian pact. I disagree.

## DEEP HISTORY

Proposals were made in the 1980s to extract mineral ores from hydrothermal vents off the coast of Oregon. But questions about technical and economic feasibility held up sea-floor mining for more than two decades. During this interval, advances in undersea technology — scientific and industrial — have yielded increasing access to the deep sea.

At first it was scientists — American and French — who dominated deep-sea exploration and research, ranging the ocean depths in their manned submersibles *Alvin* and *Nautilie*. Next, the oil and gas industry pushed into deeper and deeper waters, facilitated by advances in off-shore capabilities, and sometimes highlighting new risks and regulatory limitations, as seen in the Deepwater Horizon oil spill. Today many nations — including China, with its *Jiaolong* submersible that made its maiden dive to the bottom of the South China Sea in 2010 — operate state-of-the-art deep-sea submersibles for scientific research.

So the technology for seabed mining has matured, and the metals are definitely there: deposits of copper-, zinc-, silver- and gold-rich ores have been identified at deep-sea vents in regions with moderate seas and close to onshore mining infrastructure<sup>2</sup>. At least two mining companies (Bluewater Metals of Sydney, Australia, and Nautilus Minerals) are pushing ahead with mining exploration in territorial waters of island nations in the southwest Pacific Ocean. Both companies undertook exploration expeditions late last year — Nautilus Minerals in the waters of Papua New Guinea and Bluewater Metals in Solomon Islands waters.

Last month, Nautilus Minerals was granted a 20-year mining lease by the government of Papua New Guinea for mineral extraction at a site known as Solwara 1 in the Manus Basin. The company plans to commence open-cut mining of Solwara 1 within the next few years, removing mineral ores (and organisms) to an estimated depth of 20–30 metres over an area equivalent to about 10 football fields. And in July 2011, the International Seabed Authority (ISA), which has jurisdiction over mineral resources in international waters, will review the first lease applications for exploration of sea-floor deposits on mid-ocean ridges. The China Ocean Mineral Resources Research and Development Association submitted an application last May for exploration of the Southwest Indian ridge, and in late December, Russia submitted an application for exploration work on the Mid-Atlantic Ridge. Mineral exploitation will not be limited to territorial waters.

National and international policies for conservation have not kept pace with mineral exploration and plans for extraction. Papua New

Guinea's national environment agency, for example, has not yet set aside sea-floor vent ecosystems for conservation in any systematic manner that might protect biodiversity from the effects of mining.

Policies regulating human activities in the deep sea can be arbitrary and inconsistent. As a case in point, the Food and Agriculture Organization of the United Nations lists hydrothermal vents as vulnerable marine ecosystems to be protected from regulated fishing. As a result, seamounts in the South Pacific are protected against bottom fishing. But mineral extraction, which has the potential to destroy the very same habitat, is not prohibited.

In territorial waters, standard-setting develops in an ad-hoc manner. For example, Nautilus Minerals is working in partnership with scientists to establish effective environmental guidelines and collect baseline data<sup>3</sup>. In some cases, the sampling and analysis have been more robust than those undertaken on academic research expeditions. Once mining begins, scientists may participate in monitoring and testing strategies for assessing and mitigating the impacts of mining.

As part of its mitigation plans, Nautilus Minerals has set aside a temporary reserve area of similar size and character to Solwara 1 to serve as a possible source for natural repopulation of the mine site. Under the terms of its permit from the government of Papua New Guinea, Nautilus Minerals is obliged to meet commitments for impact mitigation and restoration, and responsible mine closure.

In international waters, there are gaps — some say chasms — with regard to regulation, governance and conservation of special habitats in the deep sea, whether they are hydrothermal vents, cold seeps or deep-water coral reefs<sup>4</sup>. Instead of leaving it to chance or to the goodwill of a few companies, conservation policies should become an integral part of international seabed regulation — before the ISA grants the first exploration and mining licences. This 15-year-old agency is also responsible for establishing environmental regulations to protect the marine environment from harmful effects that might arise during resource extraction. Some have suggested that lodging leasing and environmental responsibilities in the same agency is akin to setting the wolf to guard the sheep. There can be environmental oversight at the ISA through its Legal and Technical Council, which serves as an independent advisory body, and nations with strong conservation interests can and should ensure that the actions of the ISA take into account conservation objectives.

As one of the first steps, the International Marine Minerals Society (IMMS) presented a Code for Environmental Management of Marine Mining ([go.nature.com/mte4gq](http://go.nature.com/mte4gq)) to the ISA in April 2010. According to the ISA, the code, which was an initiative of Nautilus Minerals, is “likely to serve as a model for legally binding legislation on marine mining”.

The IMMS code offers wide-ranging environmental policies for the management of commercial mining activities. Yet it falls short of providing a comprehensive conservation policy that would systematically protect natural diversity, and ecosystem structure, function and resilience, while enabling rational use. The California Marine Life Protection Act is an example of one such effort to engage stakeholders, scientists, resource managers and members of the public in increasing the coherence and effectiveness of the state's marine management through the design of Marine Protected Areas<sup>5</sup>.

Also last year, multiple stakeholders, with the support of the ISA, the Census of Marine Life and other agencies, developed guidelines for networks of reserves for chemosynthetic ecosystems, including deep-sea hydrothermal vents<sup>6</sup>. These, or similar guidelines, need to be turned into regulations within the ISA or another competent body. Until these are in place, wholesale mining of hydrothermal vents is premature.

## UNFINISHED BUSINESS

There are three scientific reasons for deferring wholesale commercial mining until proper conservation plans are enacted. First, there is much more to learn about hydrothermal vent systems. After three decades of work, researchers continue to find new vent sites in remote locations and new species, adaptations, behaviours and microhabitats, even in well-known settings.

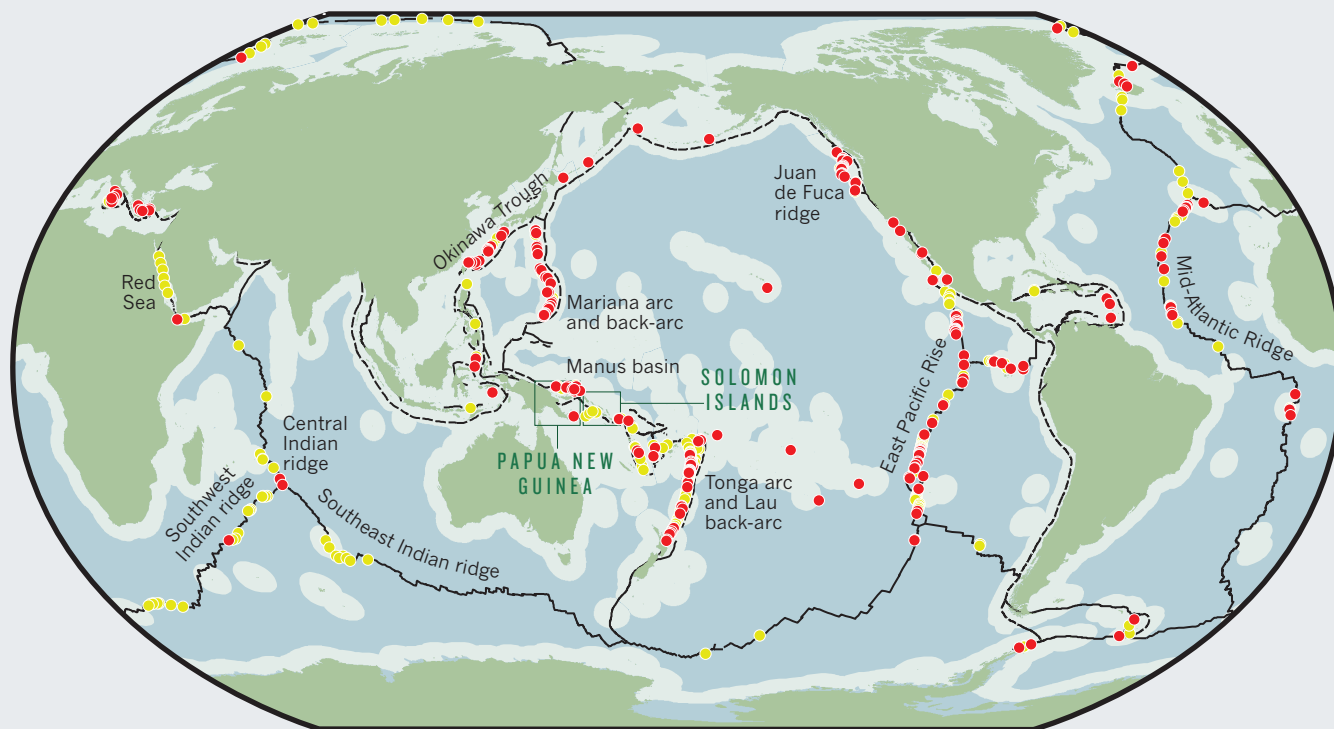
**“Deposits of copper-, zinc-, silver- and gold-rich ores have been identified at deep-sea vents.”**



## GLOBAL DISTRIBUTION OF HYDROTHERMAL VENT FIELDS

Most deep-sea vents are in volcanically active areas. Many are found in international waters, or in seas belonging to countries that are still developing deep-sea conservation policies.

● Active ● Unconfirmed — Ridge -- Trench □ Exclusive economic zones ■ International waters



S. BEAULIEU, K. JOYCE, S. A. SOULE, WHOI (2010)

Second, there is no strategy in place to assess the cumulative impacts of mining. Mining one vent field may be comparable to a volcanic eruption or other natural process that wipes out vent communities. Active hydrothermal vents are subject to frequent disturbance, including collapse of black smoker chimneys and microearthquake activity. The ability of a vent community to recover from such events may depend on their frequency as well as their scale. Moreover, scientists do not yet understand how vent systems repopulate, or anything about the complex dynamics of neighbouring communities. The effect of continuous and cumulative mining operations may be very different from that of a single event.

Third, we still don't know how best to mitigate mining activities or to restore habitats in the deep sea. Efforts by mining companies (such as setting aside a reserve area) during and after extraction could conceivably alleviate scientific concerns about cumulative effects. But which measures will work, and be affordable, won't be known until the mining is complete or until experimental studies are done.

At this point, I believe a scientific panel would review the current knowledge base and mining plans for Solwara 1 favourably — with the advice that no further mining be initiated until ecologists understand how quickly the mined vent ecosystem recovers and whether the restoration strategies used by the mining company facilitated recovery.

Marine research demands patience; expeditions are long and costly, and scientific answers slow in coming. However, we cannot be patient about effective policies to protect the sea floor. There is an urgent need to establish conservation guidelines before mining begins in international waters, and to place these guidelines in functioning governance and regulatory frameworks. Mining codes alone are not enough.

In states where seabed exploration is already under way, government agencies should act now to comply with global conservation targets, such as those adopted by the Convention on Biological Diversity. The convention has established scientific criteria to identify ocean areas

that require enhanced protection, including hydrothermal vents. It has called for a global network of comprehensive, representative and effectively managed protected areas by 2012 and suggests that at least 10% of each of the world's ecological regions be conserved. There is thus an international agreement to protect seabed vent ecosystems.

It is easy to see what would have been lost had Yellowstone been turned over to miners instead of park rangers. Kilometres of overlying water make it harder to see what would be lost in the deep sea. There are creatures of extraordinary beauty down there, exquisitely adapted to their environment. Humans may choose to threaten these habitats for economic or strategic advantage, and to feed lifestyles that depend on relentless demand for minerals and other resources. But we should make these choices on the basis of an understanding of what we may lose as well as what we may gain. ■

**Cindy Lee Van Dover**, *Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University, Beaufort, North Carolina 28516, USA.*  
e-mail: clv3@duke.edu

**The author declares competing financial interests:** details accompany this article online at [go.nature.com/b3ydzj](http://go.nature.com/b3ydzj).

1. Spiess, F. N. *et al.* *Science* **207**, 1421–1433 (1980).
2. Jankowski, P. *Independent Technical Assessment of Sea Floor Massive Sulphide Exploration Tenements in Papua New Guinea, Fiji and Tonga* (SRK Consulting, 2007); available at [go.nature.com/zffswv](http://go.nature.com/zffswv)
3. Coffey Natural Systems *Environmental Impact Statement: Solwara 1 Project* (Nautilus Minerals, 2008); available at [go.nature.com/tpkymm](http://go.nature.com/tpkymm)
4. Gjerde, K. M. *et al.* *Regulatory and Governance Gaps in the International Regime for the Conservation and Sustainable Use of Marine Biodiversity in Areas Beyond National Jurisdiction*. Marine Series I (IUCN/Natural Resources, 2008).
5. Gleason, M. *et al.* *Ocean Coast. Mgmt* **53**, 52–68 (2010).
6. Van Dover, C. L. *et al.* *Environmental Management of Deep-Sea Chemosynthetic Ecosystems: Justification of and Considerations for a Spatially Based Approach*. ISA Technical Study No. 8 (International Seabed Authority, 2011).



Al Tombaugh, the son of Pluto's discoverer, was among those defending its planetary status in 2006.

## ASTRONOMY

# The planet that never was

Neil deGrasse Tyson enjoys a passionate and personal account of the demotion of Pluto.

It is not often that people assert they have committed a capital crime. The confessed 'felon' in this case is astronomer Michael Brown of the California Institute of Technology in Pasadena. In *How I Killed Pluto and Why It Had It Coming*, he describes how his discovery of another icy body challenged Pluto's status as a planet.

The first seed of Pluto's undoing was sown on 13 March 1930, the day its discovery was announced by the Lowell Observatory in Flagstaff, Arizona. Clyde Tombaugh, a Kansas farm-boy-turned-astronomer, claimed to have found the mythical Planet X predicted by his predecessor Percival Lowell. Or had he?

If you are looking for something already named Planet X and you find it, then surely that object will get classified as a planet. And

thus, Pluto was first reported to be at least as large and as massive as Earth. This was based not on a measurement, but on the assumption that Pluto's gravity was sufficient to perturb Neptune's orbit in exactly the ways observed.

Pluto orbits sufficiently far from the Sun that all but the most advanced telescopes on Earth see it as a smudge of light. One way to measure its size is by occultation: from the time it takes Pluto to cross our line of sight to a distant star, astronomers can easily derive a lower limit for how big Pluto must be.

The problem was, every attempt to measure Pluto's shadow showed no dimming of the background star. Either Pluto was

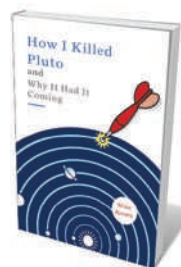
strangely transparent or it was much smaller than thought. Estimates for the planet's size kept shrinking over the decades, with a rapidity that led some wags to extrapolate (complete with a fitting function and graph) the exact date at which Pluto would disappear completely.

Pluto's size was not settled until the late 1970s, when our view of its orbit allowed James Christy of the US Naval Observatory in Flagstaff to discover its moon, Charon. From orbits and eclipse measurements, planetary scientist Richard Binzel and his collaborators discerned the masses and sizes of both bodies. Pluto was tiny. An orb measuring 2,300 kilometres across, it was not only much smaller than Earth, it was smaller than seven moons in the Solar System, with a mass equivalent to a mere 17% of our Moon.

Puny Pluto never justified its designation as the mythical Planet X. Had its true size been known to Tombaugh, he might never have classified it as a planet. The very data that led to its discovery were also challenged in 1993, when E. Myles Standish found from a re-analysis of archival observing logs at the US Naval Observatory that nothing was wrong with Neptune's orbit all along. There was no need for a perturbing Planet X. So what was Pluto?

Size alone should not be a principal factor in defining a planet — if it were, Earth could lose out too. For instance, Jupiter is 11 times wider than Earth, but Earth is only 5.5 times wider than Pluto. So Jovians might think Earth puny. For many astronomers, physical and orbital characteristics matter greatly, but in the hearts and minds of some, size still matters. That is where Brown enters the picture. In *How I Killed Pluto and Why It Had It Coming*, he explains where Pluto sits in the hierarchy of the Solar System and whether it deserves to be called a planet at all.

We learn about the 1992 discovery by astronomers David Jewitt and Jane Luu of the Kuiper belt that lies beyond Neptune — thousands of small, frozen objects that never coalesced from gravitational attraction. Pluto lives among them. Brown reasoned, as others had before, that if Pluto is large among these icy bodies yet orbits on the inner edge of the Kuiper belt, then larger objects could exist farther out that escape detection. He set out to find them. With the telescopic muscle of Caltech's observatories and plenty of youthful ambition, Brown was poised to make the discovery that transformed his life and the fate of Pluto.



**How I Killed Pluto and Why It Had It Coming**

MIKE BROWN

Spiegel and Grau:  
2010. 288 pp. \$25

**NATURE.COM**  
For a fictional take on the discovery of Pluto, see:  
[go.nature.com/xj483I](http://go.nature.com/xj483I)



Part memoir and part planetary saga, Brown's book invites you into his office, his home and his head. The account of his hard work, long hours and lost sleep reveal a dedicated researcher on a mission. He reflects on love and passion, including a charming account of how he met, courted and married his spouse. We learn about the birth of his daughter and how these domestic elements pierce his life as a scientist.

Brown's confessed crime is his 2005 discovery of Eris, an icy Kuiper-belt object that, by early estimates, was slightly larger than Pluto. What should we call it? If Eris is not a planet then it must drag Pluto down with it into the ranks of non-planethood. If we call it a planet, then Brown becomes one of only four people to have discovered one. Even he is too modest to claim that his name should hang alongside William Herschel, discoverer of Uranus, or Johann Gottfried Galle, discoverer of Neptune.

Actually, Pluto's planet status had been percolating for years. Diminutive size was only one of many factors in its demotion. Pluto's oddly tipped, elongated orbit and its icy constitution also raised eyebrows. With the discovery of the Kuiper belt, the need for an official decision grew urgent. In August 2006, at the triennial meeting of the International Astronomical Union (IAU) in Prague, a formal vote was taken on the definition of a planet.

What emerged was simple yet devastating to Pluto-lovers. Does the body mainly orbit the Sun? Is it large enough to pull its own mass into a sphere? Is its gravity strong enough for it to have (mostly) cleared its orbit of debris? Answer yes to all three and it's a planet. Given the known existence of the Kuiper belt, Pluto (and Eris) would fail the debris-free orbit criterion. And so a new term was invented for round objects that orbit in crowded places: dwarf planet.

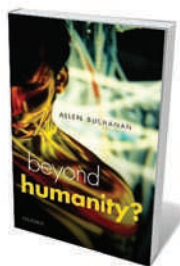
Measurements of Eris's size from a November 2010 occultation may leave Eris slightly smaller than Pluto, instead of slightly larger as Brown had previously determined. Although this revelation has resurrected the efforts of some Pluto defenders, the IAU definition remains robust against arguments of size.

So although Brown did not kill Pluto all by himself, he is guilty of providing wood and nails to construct its coffin. And my museum colleagues and I have someone to whom we can forward the hate mail we still get from Pluto-loving schoolchildren. ■

**Neil deGrasse Tyson** is an astrophysicist at the American Museum of Natural History, New York, USA. He is author of *The Pluto Files* and host of a PBS NOVA television programme of the same name.

Further reading accompanies this article online at [go.nature.com/c9gk9](http://go.nature.com/c9gk9).

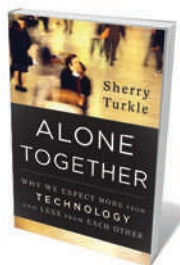
## Books in brief



### **Beyond Humanity?: The Ethics of Biomedical Enhancement**

Allen Buchanan OXFORD UNIVERSITY PRESS 256 pp. \$25 (2011)

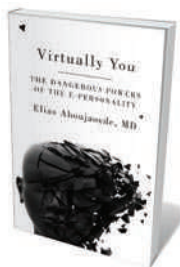
Since humans developed tools, we have sought to improve our performance through technology. Enhancements using biotechnologies should be seen in the same evolutionary context, argues philosopher Allen Buchanan. Increasing our memory, cognitive power, stamina or resistance to disease using drugs and genetic editing offers sufficient benefits to our species that we should set aside objections. He urges that evolutionary biology should be included in ethical debates about biotechnology and enhancement.



### **Alone Together: Why We Expect More from Technology and Less from Each Other**

Sherry Turkle BASIC BOOKS 384 pp. \$28.95 (2011)

The illusion of companionship fostered by technology is the focus of sociologist Sherry Turkle's latest book. From Facebook to robots, she examines how social networks give us 'friends' without the demands of intimacy, and how virtual environments allow us to overcome risk without consequences. Despite taking increasing hold of our lives, she argues, computers and robots will ultimately result in isolation, reduced privacy and diminished social skills. Yet she hopes that, by asking new questions, the young will overcome these downsides.



### **Virtually You: The Dangerous Powers of the E-Personality**

Elias Aboujaoude W. W. NORTON 349 pp. \$26.95 (2011)

Just as the persona we present to our work colleagues and our family differs, psychiatrist Elias Aboujaoude argues that we show a separate character online. From studying patients who have become mentally disturbed through excessive Internet use, he examines the construction of this e-personality, which reveals itself in the style of our e-mails, the users we associate with in our social networks and our online shopping habits. The impatient, urgent and unfocused nature of Internet usage also seeps into our offline world, he argues.



### **World Wide Mind: The Coming Integration of Humanity, Machines, and the Internet**

Michael Chorost FREE PRESS 256 pp. \$26 (2011)

Having relied since 2001 on bionic ear implants for his hearing, science writer Michael Chorost offers a personal account of the borderline between humans and machines. After exploring the technologies that might be used to fix or enhance our bodies, with a focus on brain implants, he argues that such technologies need not depersonalize us. As well as overcoming physical problems, embedded brain chips might one day transform human communication by literally plugging us into the World Wide Web.



### **Kingpin: How One Hacker Took Over the Billion-Dollar Cybercrime Underground**

Kevin Poulsen CROWN 288 pp. \$25 (2011)

Hacker-turned-journalist Kevin Poulsen investigates cybercrime in his latest book. He spotlights a notorious figure who took over a giant online criminal network and siphoned off millions of dollars from the US economy. Sought by the FBI worldwide, the hacker turned out to be security consultant Max Butler. Poulsen portrays both sides of the story and exposes the range of ongoing frauds, from phishing to Trojan viruses to counterfeiting.



## HISTORY

# The plantswoman who dressed as a boy

The tale of the first female to sail around the world deserves a more accurate telling, says **Sandra Knapp**.

In 1737, the great Swedish botanist Carl Linnaeus, who established the system by which we name animals and plants today, posed in an authentic Sami costume from Lapland not realizing that it was a woman's outfit. Another case of eighteenth-century botanical cross-dressing is related in Glynis Ridley's book. French botanist Jeanne Baret dressed as a boy to gain passage on explorer Louis-Antoine de Bougainville's voyage to circumnavigate the globe.

The intrepid Baret saw more biodiversity than the notoriously stay-at-home Linnaeus, yet she is not well known. In *The Discovery of Jeanne Baret*, Ridley purports to resurrect her name and accomplishments. Sadly, the author does not convincingly deploy the few facts available, so the book feels more like fiction than non-fiction, a novel whose characters are real people from history.

Explorers in the eighteenth century were overwhelmingly male, so the tale of a young woman who dressed as a man to see the world has huge appeal. De Bougainville set off in 1766 on the first French circumnavigation of the globe — also the first by any nation to include a professional naturalist to record the plants and animals of new lands.

The naturalist was Philibert Commerson (or Commerçon), a friend of Voltaire and correspondent of Linnaeus. He was accompanied by Baret, a peasant woman from Burgundy who was his housekeeper and lover. Ridley weaves a tale of Baret's early life from wafer-thin evidence comprising a few official documents, such as birth certificates, and a book of medicinal plants attributed to Commerson, which Ridley speculates was written by Baret. She is portrayed as a 'herb-woman' whose practical knowledge was useful to Commerson — his teacher rather than his assistant.

That Baret had the guts and determination to overcome gender and class barriers is without doubt. But Ridley does both Baret and Commerson a disservice by painting the latter as a parasitic ne'er-do-well while Baret comes across as a hard-working servant. I feel that Ridley misrepresents what must have been a truly collaborative partnership in the discovery of botanical and



**The Discovery of Jeanne Baret: A Story of Science, the High Seas and the First Woman to Circumnavigate the Globe**

GLYNIS RIDLEY  
Crown: 2010.  
304 pp. \$25

zoological wonders. Why else would Baret have stayed with Commerson in Mauritius and Madagascar after they left the expedition, only parting from him at his death in 1773?

How Baret concealed her gender for so long in the close confines of the ship is unclear. She must have bound her breasts, and Ridley's version has her hiding out in Commerson's cabin. Women disguising themselves

as men was not unknown at the time — in 1745, Hannah Snell dressed as a man and enlisted in the British Marines, serving for five years and completing tours of India; her book *The Female Soldier*, published in 1750, was hugely popular.

De Bougainville's official accounts of the voyage place the unmasking of Baret's gender in Tahiti, where the expedition stayed in April and May of 1767. Ridley, however, consulted previously unused journals and official documents and presents a different tale — predatory, sex-mad sailors and a lone vulnerable woman left unprotected by her partner. The journal kept by the ship's surgeon François Vivès is dripping with sexual innuendo; his account of the discovery of Baret's true gender is corroborated by other diaries. Ridley weaves a story of Baret's violation and subsequent pregnancy from his coy references to "*concha veneris*" (the Venus shell) and to the "Jeanneton" of French folk-song who loves her attackers. Vivès' account puts Baret's unmasking and violation on the shores of Papua New Guinea in June.

This makes compelling reading, but I would be more inclined to take the history seriously if Ridley had not got the scientific aspects so wrong. It is easy to check how many plants Commerson and Baret collected, and from where. Although Commerson never made it back to France before his death in Mauritius, the collections did. A search under

his name in the database of plants held at the National Museum of Natural History in Paris finds 1,735 specimens and their countries of origin: 234 from Madagascar, 144 from Mauritius, 84 from Brazil and so on. Most have on them old, handwritten labels with short descriptions, perhaps jotted in the field.

Ridley maintains, for instance, that Commerson and another member of the expedition, the Prince de Nassau-Siegen, collected no plants during a short stay on Java. The database differs — 50 specimens were retrieved, many of them forming the basis for new species. To highlight such inaccuracies might seem pedantic, but Ridley's story revolves around Baret struggling with heavy loads of plant presses, vials, nets and jars while Commerson swans around.

I, too, collect plants and carry paraphernalia — heavy work, but great fun. Commerson and Baret encountered a huge wealth of diversity. Commerson wrote of Madagascar:



Botanist Jeanne Baret depicted *en travesti*.

COURTESY STATE LIB. NSW, ML 980/CO1/22A2, A2327004

"I can announce to naturalists that this is the true Promised Land. Here nature created a special sanctuary where she seems to have withdrawn to experiment with designs different from those used anywhere else. At every step one finds more remarkable and marvellous forms of life."

Ridley describes how Baret discovered *Bougainvillea* in the forests of Rio de Janeiro using the doctrine of signatures, a medieval method by which herbalists attributed curative powers to plants on the basis of their appearance — a walnut was good for brain trouble, red things for wounds. Commerson had an ulcerating sore on his leg, so Ridley writes of Baret searching frantically for a cure, only to find it in the red bracts of *Bougainvillea* holding a pea-like pod that reminded her of the red-flowered runner beans from home. However, *Bougainvillea* does not have fruits like a pea, nor do the notes on the specimen mention any medicinal value.

In his notes, Commerson honoured high-ranking members of the expedition — its leader is commemorated in *Bougainvillea*, Nassau-Siegen in *Nassauvia* — but he did not publish these names. *Bougainvillea* was formally described in 1789 by Antoine Laurent de Jussieu, a French botanist who used Commerson's specimens and notes. Commerson also proposed the name *Baretia* for a Malagasy tree.

Ridley maintains that Commerson was an arrogant man who named things for himself. Yet the International Plant Names Index shows 119 species of flowering plants named in his honour — by others. None is noted as 'commersonii' on its original label. Commerson's *Baretia* was never published, not because someone wanted to do Baret down, but because it was found, on the specimen's return to Paris, that the genus already had a name.

After Commerson died, Baret married a French officer, Jean Duberna, on Mauritius and returned to France in 1774. She was awarded a state pension from 1785 in recognition of her bravery and contributions. She was not forgotten, although she never practised botany again.

Science was as collaborative then as it is now, but women's contributions were often overlooked in favour of those of male colleagues — a trend that continues today. Baret and other neglected contributors deserve recognition, but she does not need to be cast as a victim to be seen as a success, or her undoubted accomplishments overinflated. She, and women scientists in general, deserve better. ■

**Sandra Knapp** is a botanist at the Natural History Museum, London SW7 5BD, UK.  
e-mail: s.knapp@nhm.ac.uk



The mind's ability to adapt suggests that it can cope with our wired world — for better or worse.

#### NEUROSCIENCE

## Browsing and the brain

Two books reach opposite verdicts on how the Internet affects us, find **Daphne Bavelier** and **C. Shawn Green**.

Whenever a new technology reaches a tipping point of popularity, questions soon follow about its effects on society. The rise of the Internet has provoked two books probing its impact on the human brain. The fact that the authors reach opposite conclusions, despite relying on the same scientific evidence, underscores how little research has been done on this topic.

Nicholas Carr's *The Shallows* laments the possibility that long-term Internet exposure will sap us of our capacity for contemplation. At the base of his argument is the fact that the human brain is remarkably plastic. Carr makes this point compellingly using a mixture of historical anecdotes and interviews with experts in the neuroplasticity field, such as Michael Merzenich and Eric Kandel.

Having established that brains are constantly reshaped by experience, Carr argues that changes induced by Internet use, such as greater brain activation during web browsing, may not be in our best interests. If the brain adapts completely to the frenetic nature of the Internet, he warns, we may lose our capacity for absorbing practices such as reading a book. He worries that we may lose the very essence of what makes us human.

Nick Bilton's *I Live in the Future* is much more optimistic. Humming with enthusiasm for the continuing Internet revolution, he argues that social and cognitive changes are an inevitable consequence of any major technological advance and that our new abilities cannot be put back in the box.

**The Shallows: What the Internet is Doing to Our Brains/How the Internet is Changing the Way We Think, Read and Remember**

NICHOLAS CARR

W. W. Norton/Atlantic Books: 2010. 276 pp./384 pp.  
\$26.95/£17.99

**I Live in the Future & Here's How it Works: Why Your World, Work, and Brain are Being Creatively Disrupted**

NICK BILTON

Crown: 2010. 304 pp. \$25, £16.99

Such tension is to be expected whenever new forces enter society. By analogy, Carr discusses historical fears that the written word would act as a replacement for memory, resulting in humans that were 'shallower thinkers'. Bilton notes early worries that the freedom of travel offered by the railway would result in weakening moral standards. Both books review suspicions that most people would prefer to listen to a book than to read one, leading to concerns that the invention of the phonograph would kill the art of writing.

Is the Internet different? Bilton and Carr rely on the same scientific facts to argue persuasively for opposite positions. For example, functional magnetic resonance imaging studies show that Internet searches activate a larger network of brain areas than does simple text reading. Web browsing also requires additional types of mental processing — evaluating hyperlinks to make navigational decisions and filtering photos, videos and menus. As a result, brain activation is greater during Internet searches in people who are



'net savvy' than in those who are 'net naive'.

These findings cannot answer the question of whether such changes are good or bad. Conclusions are coloured by the authors' values. Bilton treats the adaption of the 'net savvy' as positive: "the brains were learning, benefiting from practice and experience". Carr comes to the opposite conclusion: "When it comes to the firing of our neurons, it's a mistake to assume that more is better."

Part of the problem is the paucity of scientific studies on the effects of modern technologies on the brain. It is a testament to both authors' skills that they were able to produce entire books on works so sparse. Unfortunately, to fill the pages, they lump information into categories that are too diverse to be useful. For example, both treat the use of all Internet technology — web browsing, web searching, texting, tweeting, video games and so on — as a single activity, despite the fact that such variety is unlikely to have one distinct effect. As with food, the effects of technology will depend on what type of technology is consumed, how much and for how long.

History suggests that technology does not change the brain's fundamental abilities. The general principles of brain organization have not changed for thousands of years — probably since the rise of language. Major technological advances do not create *de novo* brain structures. They do, however, take advantage of the cognitive flexibility of the human mind.

With each new technological development, we see a shift in the cognitive abilities and brain functions that society values most. The advent of writing systems, so celebrated by Carr, devalued the role of oral memorization through storytelling as cherished by the Greeks. Great orators such as Socrates would have lamented that Carr has lost the memory skills necessary for passing on knowledge through stories to future generations. Yet he has gained other skills by entraining alternate brain networks for reading and text analysis.

Just as it was difficult to say at the time whether the advent of writing was good or bad, a value judgement of the effect of the Internet is impossible. But it is a tribute to neural plasticity that, with each new technological development, our brains adapt — for better or for worse. ■

**Daphne Bavelier** is a professor in the Department of Brain and Cognitive Science at the University of Rochester, New York 14627, USA.  
e-mail: daphne@cvs.rochester.edu

**C. Shawn Green** is a cognitive scientist in the Department of Psychology at the University of Minnesota, Minneapolis, Minnesota 55455, USA.

## ART

# Abstract relativity

A Paris exhibition contrasts 1920s depictions of the fourth dimension, find **Stefan Michalowski** and **Georgia Smith**.

The birth of modern physics a century ago fired artistic as well as scientific imaginations. This can be seen in the Pompidou Centre's current exhibition of abstract art, covering Dutch painter Piet Mondrian and the De Stijl group, led by another Dutchman, Theo van Doesburg.

A series of canvasses illustrates the evolution of abstract techniques, from the soft contours of impressionism to the spare geometry of cubism. "We arrive at a portrayal of other things, such as the laws governing matter," Mondrian wrote. Cubist techniques were inspired, in part, by the multi-dimensional mathematics of Henri Poincaré and his contemporaries.

Most of the exhibition is rightly devoted to Mondrian and the development of his recognizable mature style. From a minimal toolbox of visual elements — white canvas, black lines and simple blocks of red, yellow or blue — emerge geometric compositions of startling intensity and elegance.

Mondrian was deeply influenced by theosophy, a spiritual movement grounded in ancient texts that was bent on uncovering universal truths in art, religion and science. He penned reams of theory as to why his abstract style was the appropriate expression of these "great generalities" for modern times.

A quiet introvert from a Calvinist family, Mondrian became a mentor to van Doesburg, by contrast a flamboyant young painter who had three wives and many artistic cliques in his short life (he died aged 47). When van Doesburg moved to Paris in 1923, the two men worked closely: their canvasses form a dialogue as each sparked fresh innovations from the other. But their intense relationship exploded a year later — and one of the flashpoints was the theory of relativity.

The public learned about Albert Einstein's discoveries after the First World War, when the solar eclipse of 1919 confirmed general relativity by showing that gravity can bend light. In Paris, space-time became a catchword in avant-garde circles. Artists from futurists to Dadaists latched on to the new ideas. Van Doesburg had already flirted with spatial geometry in

**Mondrian/De Stijl**  
Centre Pompidou,  
Paris.  
Until 21 March 2011.

four dimensions: the exhibition includes some of his tesseracts, projections on paper of four-dimensional cubes. Then, in the 1920s, he began trying to evoke time and change — four-dimensional space-time — in his paintings.

Mondrian rejected van Doesburg's attempt, and the two split over it. Symbolic of their rift was van Doesburg's use of dynamic diagonal lines, which contrasted with Mondrian's strict vertical and horizontal grids. But the quarrel went deeper than diagonals: Mondrian's doggedly developed style had become too much of a constraint for his former coterie.

The De Stijl artists wanted to remake the human environment by designing furniture, buildings and cities based on their primary-coloured, idealized structures. Van Doesburg experimented with architectural designs and films incorporating the fourth dimension. Some of these products are displayed in the exhibition, but the role of the fourth dimension is not clearly shown or explained. The artists themselves do not always seem to have grasped the

difference between a fourth dimension in space versus one in time.

As the artists tried to incorporate the new-found laws of physics in their expressions of absolute truth about the Universe, history ambushed them. Their comrades in abstraction were soon brutally dismissed by the Soviet and Nazi authorities. Einstein helped to pull the rug out from under their depictions of the 'absolute' by dissolving special relativity's neat geometries into quantum theory's fuzzy clouds of probability. But Mondrian's precise vision, with its subsumed scientific borrowings, continues to intrigue and delight. ■

**Stefan Michalowski** is a former particle physicist and executive secretary of the OECD Global Science Forum in Paris.

**Georgia Smith** is a journalist based in Paris.  
e-mails: stefanm@noos.fr; georgias@noos.fr

This article does not represent the views of the OECD.



Theo van Doesburg's use of diagonals is symbolic of his quarrel with Piet Mondrian.

T. VAN DOESBURG, SIMULTANEOUS COUNTER-POSITION (1929–30), DIGITAL IMAGE © MUS. MODERN ART, NEW YORK/SCALA, FLORENCE



# CORRESPONDENCE

## Negative results need airing too

The problem of the invisibility of negative results is underlined by the media storm over a paper supporting extrasensory perception being published in a reputable psychology journal (see *The New York Times*, 5 January 2011). Although individual reports might be statistically valid in isolation, their conclusions could still be questionable — other test results of the same hypothesis must also be taken into account.

Say a study finds no statistically favourable evidence for a hypothesis at the predetermined significance level ( $P=0.05$ , for example) and, like most with negative results, it is never published. If 19 other similar studies are conducted, then 20 independent attempts at the 0.05 significance level are, by definition, expected to give at least one hit. A positive result obtained in one of the 19 studies, viewed independently, would then be statistically valid and so support the hypothesis, and would probably be published.

Statistical corrections are routinely made for multiple testing within a study, but they are important across studies too. The difficulty lies in determining the number of parallel investigations of the same hypothesis. Perhaps different disciplinary research societies could help bring these covert experiments to light.

**Nitin Gupta, Mark Stopfer**  
NICHD, National Institutes of Health, USA.  
nitin.gupta2@nih.gov

## Think bigger for conservation

The US initiative to 'think big' about landscape-conservation cooperatives is an imaginative approach to conserving species

in the face of climate change (*Nature* 469, 131; 2011) — but thinking needs to be bigger still.

Because climate change is likely to shift entire biomes, we urge proponents to include the entire continent as a management area, with flexible borders between particular units.

We suggest that such cooperatives should collaborate with and learn from other large-scale conservation ventures, such as the International Model Forest Network — an integrated resource-management system that has operated globally since 1992 — and Natura 2000, in which different sectors and agencies are collaborating across Europe to conserve biodiversity.

Cooperation between agencies at various levels and geographical locations could then be tailored to meet particular conservation requirements.

**Malgorzata Blicharska, Grzegorz Mikusinski**  
Swedish University of Agricultural Sciences, Sweden.  
malgorzata.blicharska@slu.se

## Research type can affect citation rate

You do not mention possible confounding factors in your discussion (*Nature* 468, 1011; 2010) of the reported positive effect of first and last author geographical proximity on paper citations (K. Lee *et al.* *PLoS ONE* 5, e14279; 2010).

One is that these are biomedical research papers. This field has many different author-sequence conventions and citation cultures.

In basic research, the first author on a publication is typically a PhD student and the last author is his or her supervisor. The papers come from closely knit research groups, especially in molecular biology, and tend to have zero

distance between the first and last authors, and to be cited more frequently than clinical research papers.

By contrast, clinical research projects typically have no clear hierarchical structure among collaborators, and often apply alphabetical ordering of co-authors. Hence, the type of research could also explain the positive correlation you discuss.

The challenge in training researchers to collaborate on publications is to find a balance between face-to-face discussion and the use of new communication technologies.

**Henk Moed** Elsevier, the Netherlands.

h.moed@elsevier.com

## Controversy over GM maize in Peru

Researchers from the Peruvian National Institute for Agricultural Innovation (INIA) — which has been enforcing national and international policy on biosafety in agriculture since 1999 — have investigated claims that genetically modified maize (corn) is being farmed in the Barranca valley north of Lima (see [go.nature.com/ijkpkz](http://go.nature.com/ijkpkz)).

The INIA analysed the source and quantity of maize imports, records of seed cultivars, their genetic diversity and planting location. Samples were also tested from the Pativilca River basin — the main river in Barranca and its neighbouring valleys. These came from maize fields, local markets, a local collecting facility and seed companies that sell poultry feed.

Evidence of transgenes was discovered in only some of the poultry grain samples (full details are available in Spanish at [go.nature.com/ikgyqj](http://go.nature.com/ikgyqj)). This finding is not

surprising. Peru imports about 1.5 million tonnes of maize grain annually — mainly for animal feed — from Argentina and the United States, where genetically modified maize is widely grown.

We believe that the Barranca region today is unlikely to be a primary centre of maize diversity. However, farmers there may be growing maize hybrids and other cultivars that have seeds of foreign origin.

**Luis Fernando Rimachi Gamarra, Jorge Enrique Alcántara, Rodomiro Ortiz**  
Instituto Nacional de Innovación Agraria, Perú.  
rodmiroortiz@gmail.com

## Self-plagiarism in music and science

Composers are much more relaxed about self-plagiarism than scientists. It was practised by the best: take Bach's *Christmas Oratorio*, which recycles several of his secular cantatas, and Mozart's *Mass in C Minor*, which was transformed into his *Davidde Penitente*.

As for Handel, he was prone to reproducing his own and his colleagues' music with equal nonchalance. His love duet 'No [pause] di voi non vo' fidarmi' becomes 'For [pause] unto us a child is born' in *Messiah*. Same music, different atmosphere.

Some scientists might also defend self-plagiarism on the grounds that the data are the same but the conclusions are not. Even my venerable professor of biochemistry, when I chided him for setting his students the same exam questions he had asked us 20 years before, replied tersely, "The questions are the same, the answers are different."

**Renato Baserga** Kimmel Cancer Center, Pennsylvania, USA.  
r\_baserga@kimmelcancercenter.org

# Eugene Goldwasser

## (1922–2010)

Discoverer of the hormone that regulates the production of red blood cells.

Eugene 'Gene' Goldwasser made one of the outstanding advances in twentieth-century biomedicine. Through decades of effort, he purified and initially characterized the properties of the major hormonal regulator of red-blood-cell production — erythropoietin (EPO) — an advance that rivals the discovery of insulin in its importance. Goldwasser's work was instrumental to the large-scale production of recombinant EPO, which since the late 1980s has been used to induce red-cell formation, especially in anaemic patients with chronic renal disease.

Hundreds of thousands of people now benefit each year from this therapy, which has a multibillion-dollar annual market. Goldwasser himself — who died on 17 December 2010, from renal complications associated with prostate cancer — did not become rich. He was more captivated by the science than the money. Today, studies of EPO continue to provide breakthroughs in cytokine and receptor biology.

Goldwasser was born in Brooklyn, New York, in 1922, but the Depression forced his family to close their clothing business and move to Kansas City. In high school, Goldwasser's scientific interests were already keen, and landed him a scholarship at the University of Chicago, Illinois, for his undergraduate degree in biological sciences. Following two years of US army service as a biochemist working on anthrax at Fort Detrick in Frederick, Maryland, Gene returned to the University of Chicago and completed his PhD in biochemistry in 1950. As a postdoctoral fellow in Copenhagen he trained in cytophysiology — the study of the biochemistry of cells. In 1952 he returned to the University of Chicago. His initial studies there, at the Argonne Cancer Research Hospital, focused on new approaches to treating leukaemia.

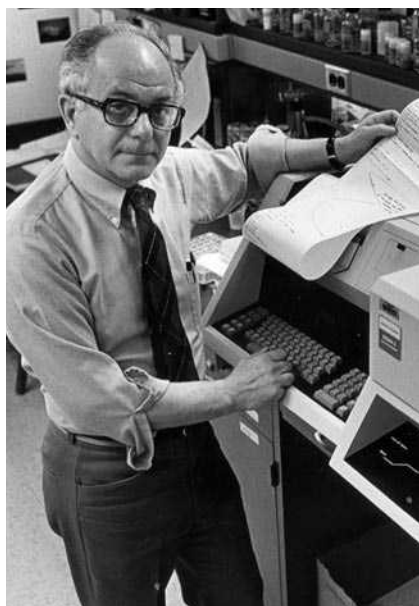
### HORMONE HUNT

The first clues to EPO's existence were found more than a century ago. This early work relied on highly variable measures of the number of mature blood cells, so the existence of such a substance remained contentious. In the 1940s, researchers had more precise measures of red-cell precursors, and named the proposed substance EPO; by the 1950s, more definitive results came from experiments using surgically conjoined rats and acutely anaemic animals. Still, no one knew exactly what was stimulating

the formation of red blood cells.

In 1953, Goldwasser turned his attention to the EPO enigma. From 1956 to 1959, he worked aggressively and published 15 EPO-related papers, at a time when not everyone was convinced that EPO even existed. These included seminal studies published in *Nature* and *Science* showing that the kidney was a prime site of EPO production, and that cobalt could induce EPO production. He also developed a sensitive assay for EPO activity using a radioactive isotope of iron, which enabled the development of purification strategies.

Goldwasser originally estimated it would take him about six months to purify EPO.



It took 17 years. EPO from any source existed in only vanishingly low quantities, and it tended to degrade during attempts at purification. In early studies, Goldwasser used litres of plasma from anaemic sheep, but even increasing the EPO concentration 1-million-fold didn't produce a pure sample. He tried many sources, and had some luck with urinary protein from anaemic Argentine patients with hookworm. A major breakthrough came on Christmas morning in 1975, when Goldwasser's colleague Takaji Miyake of Kumamoto University, Japan, hand-delivered a special gift: processed protein from more than 2,500 litres of urine from patients with aplastic anaemia. Others might have preferred fruit cake. Goldwasser and his colleagues used this unique gift to

finally purify, in 1977, about 8 milligrams of human urinary EPO.

By the 1980s, other researchers began entering the field that Goldwasser had opened up. I joined Children's Hospital Boston and Harvard Medical School in Boston, Massachusetts, to work on EPO as a postdoctoral fellow. When I met Gene for the first time at a conference, he smiled and said: "It's nice to have you in the field, but should you not be in Chicago?" — working with him, in other words.

Goldwasser was captivated by EPO's basic science, but was also keenly aware of its potential clinical importance. In 1978, for example, he was involved with an early small clinical trial. At his university, he had initiated some paperwork for an invention disclosure, but this did not move forward. In the early 1980s, the idea of commercializing such work was quite new, as were the first biotechnology companies.

Goldwasser did agree to consult for a biotechnology company — Amgen. During the 1980s, some of the peptide fragments of human EPO were sequenced — a first step towards cloning the human EPO gene and producing clinically useful quantities of recombinant protein. As testimony to his excitement, Goldwasser unwittingly presented one such sequence at a national meeting. Fortunately for Amgen, the sequence turned out to have several mistakes, and no one else gained from the information. Amgen eventually took the lead in commercializing EPO.

Gene truly had a lifelong fascination for science. Had he chosen, he could have made great contributions in other areas of biochemistry: in 1953, for example, he published impressive work in *Nature* on nucleotide biosynthesis. From 1994 to 1998, he returned from retirement to chair the University of Chicago's biochemistry department, and continued his studies of EPO from 1998 to 2002, eventually retiring again at the age of 80. His unassuming nature, critical mind, collegiality and dedication — including to the Chicago Blackhawks ice-hockey team — will be greatly missed. ■

**Don Wojchowski** is director of the COBRE Center of Excellence in Stem Cell Biology and Regenerative Medicine at the Maine Medical Center Research Institute, Scarborough, Maine 04074, USA.  
e-mail: wojchd@mmc.org

## FORUM Drug discovery

# A question of library design

Two approaches have emerged for creating libraries of compounds for use in biological screening assays for drug discovery — fragment-based ligand design and diversity-oriented synthesis. Advocates of each approach discuss their favoured strategy.

### THE TOPIC IN BRIEF

- There is an urgent need to improve the libraries of compounds used for drug discovery, to find better leads for medicinal chemistry programmes.
- Fragment-based ligand design involves screening small molecules that aren't intrinsically drug-like, but that might become subunits (fragments) of drug-like compounds.
- Diversity-oriented synthesis aims to

make many structurally varied, drug-like compounds for screening, using modular syntheses that involve few steps.

- The lead compounds identified from diversity-oriented synthesis generally differ markedly from those obtained from fragment libraries.
- The pros and cons of the two approaches, and the chances of success of subsequent drug programmes, are a matter of vigorous debate.

## Small molecules, great potential

PHILIP J. HAJDUK

Both fragment-based screening (FBS) and diversity-oriented synthesis (DOS) grew out of the recognition 15 years ago that the libraries of compounds available for high-throughput screening (Fig. 1) were inadequate for many lead-discovery campaigns. As a result, almost every large pharmaceutical company undertook library enrichment exercises, all of which incorporated some aspect of DOS<sup>1</sup>. The challenge has always been to balance the size and structural diversity of compound collections against the cost associated with screening the compounds, while addressing the needs — binding affinity, selectivity and so on — of the anticipated portfolio of biological targets. FBS and DOS represent two extreme views on how to address these issues. I believe that FBS is the better strategy.

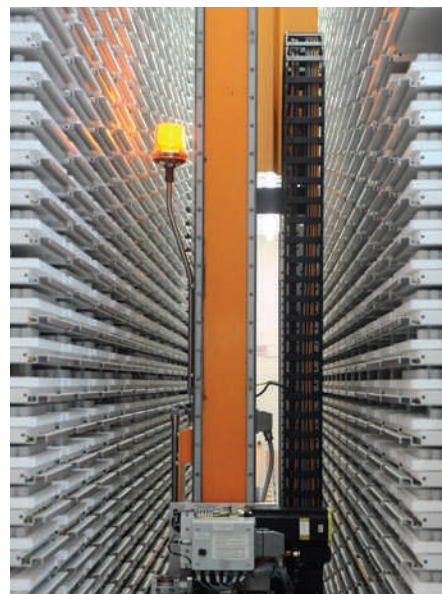
Proponents of DOS advocate the production of numerous sets of compounds that have molecular structures not represented in existing libraries, to continuously fill the gaps in 'chemical-diversity space'. But this approach potentially yields millions of compounds that must all be screened at the start of drug-discovery programmes. Proponents of FBS believe that tremendous (and probably sufficient) chemical diversity can be represented in a library of several thousand 'fragments' (Fig. 2a). Indeed, fragment collections of as few as 1,000

molecules can arguably represent the chemical diversity contained in tens of millions of larger, more drug-like compounds<sup>2</sup>. Thus, FBS libraries achieve greater chemical diversity than even the largest available compound libraries, and can be screened far more cost-effectively.

Most DOS libraries are prepared in a purely speculative manner, in the sense that it is not known if members of the libraries will be active against any relevant biological target. A significant up-front investment in compound synthesis must therefore be made that may never pay off. By contrast, compounds based on fragment leads are always directed towards, and dictated by, the target under study. This allows chemists to dedicate their efforts primarily to current drug-discovery targets, rather than diverting resources to the potentially wasteful production of compounds that have unknown biological activities.

Finally, there is ample evidence that larger molecules are less likely than smaller ones to succeed as drugs in clinical trials<sup>3</sup>, mainly because their physico-chemical properties are not drug-like. It is therefore vital to identify drug candidates that not only are potently active at a biological target, but also have acceptable physico-chemical properties. Because fragment-based design involves the tailored construction of drugs from compounds that are soluble and of low molecular mass, this strategy offers the greatest potential for discovering the smallest possible compounds that bind most efficiently to a particular target — that is, compounds in which all the structural features contribute to binding.

By contrast, many of the properties of



**Figure 1 | Stack 'em up.** Part of the compound library at the Sanofi-aventis laboratory in Toulouse, France. More than 1 million compounds, stored in trays of vials, are kept here.

compounds in DOS libraries are not drug-like. So, even if these compounds appear as hits — active compounds — in a screen, many analogues may have to be made to find one that is not only active at a biological target, but also 'druggable'. Indeed, most compounds in DOS libraries would be excluded from many corporate screening collections because of their poor physico-chemical properties.

In summary, fragment-based drug design offers several advantages over DOS: fragment libraries are more diverse, synthetic resources are used more efficiently and the leads identified from FBS are more likely to yield drug candidates that have optimal physico-chemical properties. Indeed, several compounds derived from fragment-based drug design are already in clinical trials<sup>4</sup>, providing substantial justification for further investment in this strategy by the pharmaceutical industry.

**Philip J. Hajduk** is in the Department of Lead Discovery, Abbott Laboratories, Abbott Park, Illinois 60064, USA.  
e-mail: philip.hajduk@abbott.com



# Better leads come from diversity

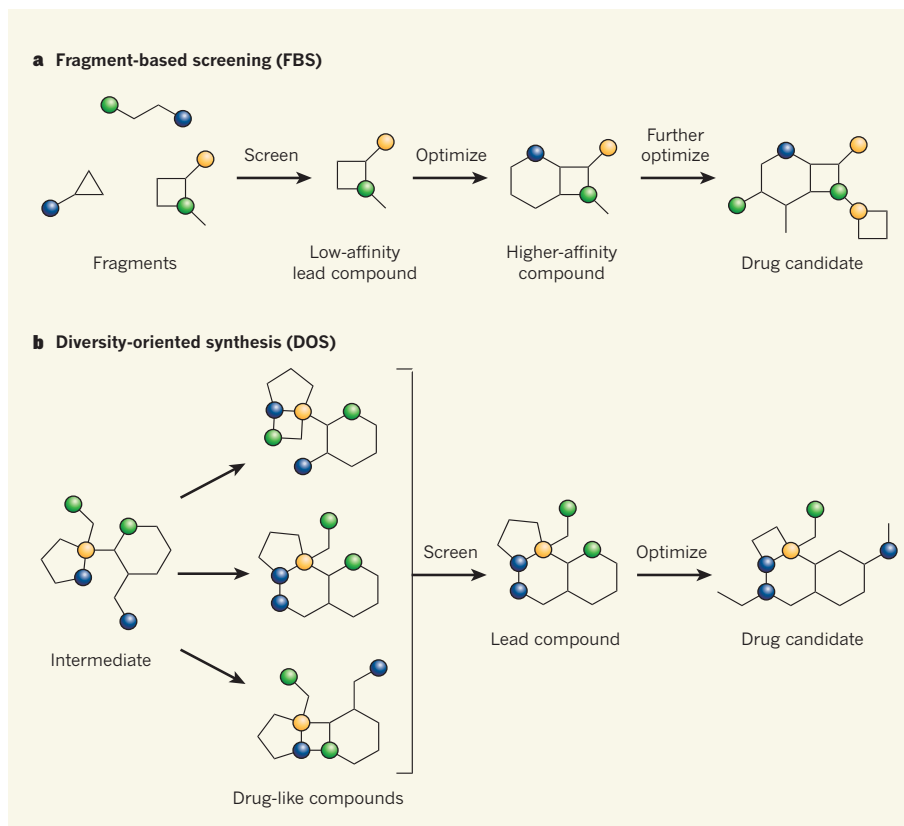
WARREN R. J. D. GALLOWAY &  
DAVID R. SPRING

The advantage of the DOS approach is that it efficiently creates structurally diverse molecules whose molecular masses are usually close to those of drug-like compounds<sup>5,6</sup> (Fig. 2b). Moreover, DOS provides access to molecules that have thus far escaped the attention of humans and perhaps even nature<sup>5</sup>. We believe that screening such molecules is the best general approach to finding a new lead compound for drug discovery. Let us illustrate why.

Put yourself in the shoes of a scientist who wants to develop a new drug to treat a disease such as cancer, but who does not know the precise nature of the relevant disease-causing target. A direct way to address this problem is to screen a library of molecules to see whether any of them kill cancer cells selectively. To maximize success, you need a collection of structurally diverse, drug-like molecules, such as those produced from DOS. A fragment library would be completely inappropriate, because molecules this small do not bind to drug targets with sufficient potency and specificity to be identified in such screens.

The use of a fragment library is viable, however, when you know exactly what the protein target is. To be fair, this is often the case for pharmaceutical companies today. But because fragment-based drug discovery requires knowledge of the way in which substrates bind to targets, this approach works only if you have a water-soluble protein for which much structural information is available, and for which the molecular binding modes of substrates are easily obtainable using methods such as nuclear magnetic resonance (NMR) spectroscopy or X-ray crystallography<sup>4,7,8</sup>. Unless this is the case, then screening structurally diverse drug-like molecules (such as those obtained using DOS) in biological assays is the way forward.

Even if structural information about a target is easy to obtain, FBS won't necessarily provide any hits. We have experienced this when looking for modulators of protein–protein interactions (PPIs). Indeed, it can be argued that non-traditional pharmaceutical targets — those, such as PPIs, that don't involve enzymes or receptors — are unlikely to be suitable for fragment-based drug discovery. This is because the binding sites that are tractable to pharmaceutical modulation in these systems are typically more highly exposed to water than are enzyme active sites or receptors; fragments tend to bind poorly to such sites. Yet such 'undruggable' targets are currently the most exciting for drug discovery. Screening



**Figure 2 | Generating lead compounds for drug discovery.** **a**, In fragment-based screening, libraries of structurally diverse, small molecules that could become fragments of active drugs are screened in assays for a biological target. The lead compounds identified in this way have low affinities for the target, but subsequent rounds of optimization — in which the structure of a lead is systematically altered and enlarged — generate high-affinity, drug-like compounds for clinical trials. Differently coloured circles represent different chemical groups. **b**, Using diversity-oriented synthesis, libraries of structurally diverse, drug-like compounds are made as efficiently as possible, typically from common intermediates. The libraries are then screened and the resulting lead compounds — which typically have higher affinities for targets than do fragments — are optimized to produce candidates for clinical trials. The drug candidates produced using the two approaches tend to differ from each other in many respects. Libraries produced by either method typically contain hundreds or thousands of compounds.

of DOS libraries has provided hits for several non-traditional targets, including PPIs<sup>9</sup>.

Of course, DOS has its own inherent challenges — fragment libraries can, in principle, cover more chemical space with fewer compounds of a given molecular size than can DOS libraries, for example. But, in practice, fragment libraries often have limited structural diversity and tend to be biased either towards compounds that satisfy the dogma of traditional medicinal chemistry or towards aromatic compounds (those containing benzene rings or related ring structures), which are easily detected by NMR screening. Experience also shows that the optimization of leads from FBS is likely to generate flat molecules as drug candidates. Nature, however, is three-dimensional, and so drugs are likely to be more selective for their targets if they too are three-dimensional. An advantage of DOS is that it typically comes up with new, three-dimensional molecular scaffolds.

To be clear, we do acknowledge that FBS has led to the discovery of drug-like compounds

in certain optimal cases. But we believe that better compounds can be found using DOS, an approach that is applicable for drug discovery in general. ■

**Warren R. J. D. Galloway and David R. Spring** are in the Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK.  
e-mail: spring@ch.cam.ac.uk

1. Jacoby, E. *et al. Curr. Top. Med. Chem.* **5**, 397–411 (2005).
2. Hajduk, P. J., Meadows, R. P. & Fesik, S. W. *Science* **278**, 497–499 (1997).
3. Leeson, P. D. & Springthorpe, B. *Nature Rev. Drug Discov.* **6**, 881–890 (2007).
4. Hajduk, P. J. & Greer, J. *Nature Rev. Drug Discov.* **6**, 211–219 (2007).
5. Galloway, W. R. J. D., Isidro-Llobet, A. & Spring, D. R. *Nature Commun.* **1**, 80, doi:10.1038/ncomms1081 (2010).
6. Schreiber, S. L. *Nature* **457**, 153–154 (2009).
7. Hajduk, P. J. *Nature Chem. Biol.* **2**, 658–659 (2006).
8. Erlanson, D. A. *Curr. Opin. Biotechnol.* **17**, 643–652 (2006).
9. Di Micco, S. *et al. J. Med. Chem.* **52**, 7856–7867 (2009).

# Defence at dawn

A remarkable example has been discovered of a plant tuning its immune defence against a pathogen. The tuning consists of maximal expression of the relevant genes at the time of day when attack is most likely. [SEE LETTER P.110](#)

C. ROBERTSON MCCLUNG

Plants cannot avoid pathogens by fleeing, so they have evolved mechanisms to resist attack. Those mechanisms include forms of immunity, and on page 110 of this issue Wang *et al.*<sup>1</sup> reveal an unexpected connection between plant immunity and the circadian clock.

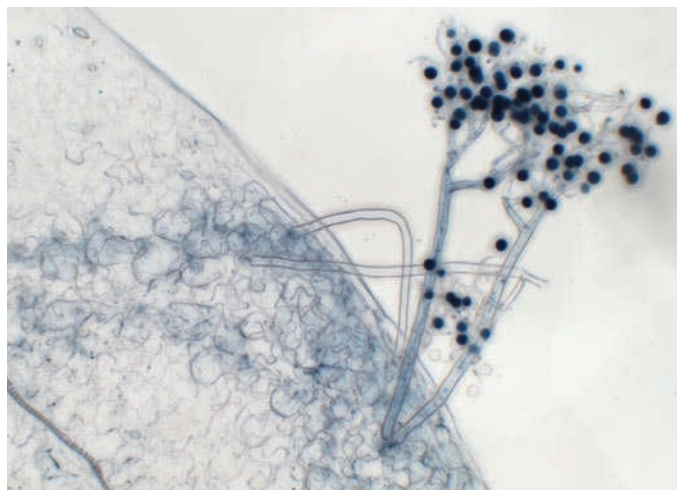
Plant immunity is controlled by a complex and incompletely understood signalling network. The first level of defence consists of the recognition of molecules called pathogen-associated molecular patterns (PAMPs) and the subsequent initiation of PAMP-triggered immunity<sup>2</sup>. Because many pathogens can suppress this response, often by means of proteins generically termed effectors, plants have evolved disease-resistance (*R*) genes that recognize effectors or the action of effectors on their host-cell targets. The consequent effector-triggered immunity involves the reprogramming of gene transcription, which culminates in physiological changes that include the hypersensitive (cell death) response<sup>2</sup>.

Plant pathogens can be divided into necrotrophs, which kill the host and feed on the dead tissues, and biotrophs, which infect and feed off a living host without killing it. *R*-gene-mediated defence is the primary response to biotrophs, including the oomycete *Hyaloperonospora arabidopsidis*, which causes downy mildew disease in the thale cress *Arabidopsis* (Fig. 1). In their paper, Wang *et al.*<sup>1</sup> describe novel components of *R*-gene-mediated resistance against downy mildew in *Arabidopsis*, and show that these components are controlled by a transcription factor called CIRCADIAN CLOCK-ASSOCIATED 1 (CCA1), which is an essential component of the circadian clock.

Wang *et al.*<sup>1</sup> employed a systems approach. They compared changes in *Arabidopsis* gene expression in response to *H. arabidopsidis* infection over time in wild-type seedlings (which successfully resist this pathogen owing to the presence of the *RPP4* disease-resistance gene) with gene-expression changes in an *rpp4* mutant (which has lost *RPP4* function and is sensitive to infection). Their examination of

events shortly after infection revealed several genes that responded to *H. arabidopsidis* challenge in an *RPP4*-dependent manner but that had not previously been identified as immune regulators or as components of the response to *R*-gene activation.

The authors then tested mutants in which each of these novel plant genes was non-functional, to define a subset that is important for successful disease resistance. These genes could be divided into two clusters — one associated



**Figure 1** | *Hyaloperonospora arabidopsidis* in spore-production mode. The spores (dark particles) are borne on a structure known as the sporangiophore, which here is about 0.7 millimetres in height.

mainly with the hypersensitive response and the other with physico-chemical responses to infection. Intriguingly, the promoter sequences of both groups of genes showed over-representation of binding sites for CCA1, implicating the circadian clock in the regulation of immunity.

The circadian clock is a timekeeping mechanism that enables the anticipation of environmental conditions or biological events that occur at predictable times of day<sup>3</sup>. For example, cold stress typically occurs at night, and the plant circadian clock modulates cold responses accordingly<sup>4</sup>. Similarly, *H. arabidopsidis* sporulates at night and disseminates its spores at dawn, perhaps because enhanced humidity optimizes germination and the colonization of new hosts<sup>5</sup>.

Wang *et al.*<sup>1</sup> found that some *Arabidopsis* mutants in which clock function was impaired showed increased susceptibility to

*H. arabidopsidis* infection. They propose that, in wild-type plants, disease resistance is maximal at dawn, when the likelihood of encountering *H. arabidopsidis* spores is greatest, and reduced at dusk. They suggest that the plant circadian clock mediates a pulse of defence-gene expression near dawn, when *CCA1* expression is high, in anticipation of possible *H. arabidopsidis* challenge at that time. This anticipatory expression of defence genes, at a time when the potential for pathogen challenge is perhaps higher than during the rest of the day, could allow the need for maximal disease resistance to be balanced against the growth decrement associated with sustained expression of defence genes<sup>6</sup>. The central role of *CCA1* is emphasized by the reduced resistance to *H. arabidopsidis* at dawn in *cca1* loss-of-function mutants; these same mutants showed no reduction in resistance, compared with wild-type plants, in response to *H. arabidopsidis* challenge at dusk, when *CCA1* is not expressed.

It is intriguing that *CCA1* serves as an integrator between the immune response and the circadian clock to elicit enhanced resistance, whereas a closely related transcription factor, *LHY*, does not — *lhy* mutants show no decrement in disease resistance, yet display a similar shortening of circadian period to that seen in *cca1* mutants<sup>3</sup>. That may indicate that the primary defect in disease resistance in *cca1* mutants is due to the action of *CCA1* on an output pathway that is modulated by the circadian clock, rather than on the oscillator mechanism itself. It is also intriguing that *CCA1*, but again not *LHY*, has a crucial role in the integration of nitrogen assimilation and the circadian clock<sup>7</sup>.

The work of Wang *et al.*<sup>1</sup> has implications for understanding other plant pathogens. One example is another oomycete, *Phytophthora infestans*, which causes late blight in potatoes and was responsible for the potato famines of the 1840s<sup>8</sup>. A second example, of more recent notoriety, is the biotrophic Ug99 strain of wheat stem rust (*Puccinia graminis*, a basidiomycetous fungus), which was first identified in Uganda in 1999 but now threatens wheat-growing regions in Asia<sup>9</sup>. Ug99 overcomes the resistance afforded by most of the *R* genes in current cultivars. New insights into *R*-gene-mediated disease resistance may contribute to the development of wheat that is resistant to this devastating pathogen.

The results<sup>1</sup> may also apply beyond the plant world. For instance, many genes involved in innate immunity in fruitflies exhibit cyclic expression controlled by the central oscillator component *CLOCK*<sup>10</sup>. Wild-type flies show circadian variation in resistance to bacterial

W. WANG & X. DONG

infection, peaking in the middle of the night<sup>11</sup>. Similarly, evidence is accumulating in support of circadian modulation of human immune responses<sup>12</sup>.

Finally, returning to *H. arabidopsidis*, it is not known whether the apparently rhythmic sporulation<sup>5</sup> is a direct response to the light-dark cycle or instead represents a circadian rhythm driven by an internal clock. Nor is it known whether rhythmic sporulation is a characteristic of oomycetes in general. Rhythmic asexual spore production (conidiation) in the ascomycetous fungus *Neurospora crassa* has provided one of the most fruitful systems for investigating circadian rhythms<sup>13</sup>, and evidence from genome sequencing of many fungal species indicates that some clock proteins were

present in the common ancestor of the fungi<sup>14</sup>. Oomycetes are only distantly related to fungi<sup>15</sup>, but the identification of a circadian component in plant resistance to oomycete infection justifies investigation of potential circadian rhythmicity in oomycetous and fungal pathogens as well<sup>16</sup>. ■

**C. Robertson McClung** is in the Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA.  
e-mail: c.robertson.mcclung@dartmouth.edu

1. Wang, W. *et al.* *Nature* **470**, 110–114 (2011).
2. Jones, J. D. G. & Dangl, J. L. *Nature* **444**, 323–329 (2006).
3. McClung, C. R. *Plant Cell* **18**, 792–803 (2006).

4. Mikkelsen, M. D. & Thomashow, M. F. *Plant J.* **60**, 328–339 (2009).
5. Slusarenko, A. J. & Schlaich, N. L. *Mol. Plant Pathol.* **4**, 159–170 (2003).
6. Heil, M. & Baldwin, I. T. *Trends Plant Sci.* **7**, 61–67 (2002).
7. Gutiérrez, R. A. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 4939–4944 (2008).
8. Reader, J. *Potato: A History of the Propitious Esculent* (Yale Univ. Press, 2009).
9. Marris, E. *Nature* **456**, 563–568 (2008).
10. McDonald, M. J. & Rosbash, M. *Cell* **107**, 567–578 (2001).
11. Lee, J.-E. & Edery, I. *Curr. Biol.* **18**, 195–199 (2008).
12. Berger, J. J. *Appl. Biomed.* **6**, 65–72 (2008).
13. Dunlap, J. C. & Loros, J. J. *Curr. Opin. Microbiol.* **9**, 579–587 (2006).
14. Salichos, L. & Rokas, A. *Mycologia* **102**, 269–278 (2010).
15. Soanes, D. M., Richards, T. A. & Talbot, N. J. *Plant Cell* **19**, 3318–3326 (2007).
16. Roden, L. C. & Ingle, R. A. *Plant Cell* **21**, 2546–2552 (2009).

## ASTROPHYSICS

# Big black hole found in tiny galaxy

**Conventional wisdom tells us that supermassive black holes are found exclusively in massive galaxies undergoing little star formation. But one such object has now been discovered in a star-forming dwarf galaxy. SEE LETTER P.66**

JENNY E. GREENE

On page 66 of this issue, Reines and collaborators<sup>1</sup> report a most unlikely discovery: a big black hole in the centre of a very small galaxy. The low-mass (or 'dwarf') galaxy Henize 2-10, the focus of their research, is currently undergoing a major growth spurt. It seems to have formed an appreciable fraction of its stars in the past 10 million years<sup>2</sup>, which is unusually fast for present-day galaxies. Now it appears that Henize 2-10 is growing more than just stars. The authors detect a source of light that they attribute to gas falling into a supermassive black hole at the heart of the galaxy. This is the first potential detection of a black hole in a rapidly star-forming dwarf galaxy such as Henize 2-10. If confirmed, this observation gives us the hitherto impossible opportunity to study at first hand the growth of black holes in a forming galaxy.

Over the past three decades, astronomers have used Henize 2-10 and galaxies like it as laboratories for what might have happened in the first days of galaxy formation<sup>3</sup>. They can study how gas is converted into stars, and how and where those stars are formed, as the galaxy grows. It turns out that most young stars form in compact clusters within the growing galaxy.

Despite many years of study, however, no central supermassive black hole had been discovered in Henize 2-10. Light from the young, newly forming stars had overwhelmed the

unique light signatures of the black hole and masked its presence. But using new and archival high-resolution data, Reines *et al.*<sup>1</sup> found a source of both radio and X-ray emission far from any site of new star formation. Although massive star birth and death is accompanied by sources of radio and X-ray emission, the relative strengths of the radio and X-ray sources, combined with their distance from newly forming stars, strongly suggests a more exotic origin — a radiating black hole.

Reines and collaborators were aided by better data, but ultimately it was good intuition and imagination that led to their discovery. Conventional wisdom tells us that supermassive black holes with masses millions to billions times that of our Sun are found exclusively in massive galaxies, not in dwarf galaxies such as Henize 2-10. And usually, supermassive black holes are found in the parts of galaxies with regular elliptical shape and very little ongoing star formation, not in lumpy star-forming galaxies. The new discovery, while unexpected, may represent an important opportunity to investigate the unknown origins of supermassive black holes. We do not yet have telescopes powerful enough to witness the interactions between the first black holes and their parent galaxies. Because Henize 2-10 is currently forming a large fraction of its stars, much like the host galaxies of the first black holes, we may finally have our chance to observe a growing black hole within a forming galaxy.

The case is not yet watertight for a black hole in the centre of Henize 2-10. We cannot rule out the chance alignment of two less exotic objects associated with star birth and death in such a complicated and dusty region. Furthermore, some expected signatures of a black hole, such as emission from highly excited neon atoms, are not observed. But without knowing more precisely the mass and total power of the putative black hole, it is hard to guess how bright the emission should be. Reines *et al.* estimate a range of black-hole masses between 100,000 and 10 million times the mass of the Sun, which places it at the low-mass end of supermassive black holes. Accurate estimates of the black-hole mass must await the next generation of optical telescopes. In the meantime, Reines and collaborators are obtaining a more precise measure of the total power output of the black hole.

Although the discovery of a massive black hole in a dwarf star-forming galaxy is unexpected, there are many known examples of black holes in small galaxies. Indeed, recent work<sup>4–8</sup> has uncovered many examples of small galaxies containing black holes with masses approximately 10,000 times that of our Sun. Unfortunately, obtaining an accurate count of the black-hole population becomes increasingly difficult for low-mass black holes in small galaxies. Emission from these black holes is always faint, and because they are so small we cannot directly observe their gravitational pull on surrounding stars.

Reines and collaborators<sup>1</sup> have identified an entirely new parent population of galaxies where many unknown supermassive black holes could be lurking. There are many things Henize 2-10 may teach us. First, black-hole growth might inhibit star formation in fragile, first-generation galaxies<sup>9</sup>. To test this hypothesis, we can look for signs that energy from the black hole is heating or removing gas in Henize 2-10. Second, the location of the putative black hole away from any star cluster is intriguing, because most proposed mechanisms for making new black holes involve massive stars<sup>10</sup>. And perhaps



most importantly, we will now start looking for more black holes in dwarf star-forming galaxies. Henize 2-10 reminds us to keep our eyes open and expect the unexpected. ■

**Jenny E. Greene** is in the Department of Astronomy, University of Texas, Austin, Texas 78712, USA.  
e-mail: jgreene@astro.as.utexas.edu

## MOLECULAR BIOLOGY

# DNA fragility put into context

**Fragile sites are genomic regions prone to deletions or other alterations during DNA replication. The reason for the susceptibility of these sites to damage may be simple: they contain few replication initiation sites. [SEE LETTER P.120](#)**

KAY HUEBNER

Agents that hinder DNA replication can often cause breaks at specific chromosome regions commonly called fragile sites. Two questions have been hotly debated concerning these chromosome regions: why are they so sensitive to DNA damage; and how does damage to genes encompassing such sites affect the biology of the damaged cells? A paper by Letessier *et al.*<sup>1</sup> on page 120 of this issue presents a rather simple, yet elegant, answer to the first of these questions. The authors also report that, surprisingly, the sensitivity of specific fragile sites within a cell depends on the tissue or organ from which the cell originates — a conclusion that may provide clues to the answer to the second question.

During DNA replication, the enzyme helicase breaks hydrogen bonds that hold the two DNA strands together, uncoiling the DNA spiral to form a structure called the replication fork. In humans, mice and other mammals, the propensity of common fragile sites to deletions and other rearrangements has been attributed to various features of the specific DNA sequences at fragile regions, which can span more than 1 million base pairs. For instance, DNA sequences at fragile sites might be prone to forming secondary structures that impair the movement of the replication fork, leading to its collapse and to DNA breaks<sup>2</sup>; fragile regions might replicate late in the DNA-replicative phase — even later in the presence of DNA-damaging agents — and so their replication remains incomplete<sup>3</sup>; or differences in the composition of DNA-associated proteins between fragile and non-fragile regions<sup>4</sup> may account for the susceptibility of the former to breakage.

Letessier and colleagues<sup>1</sup> point to a different reason. The authors focus on *FRA3B* — the

1. Reines, A. E., Sivakoff, G. R., Johnson, K. E. & Brogan, C. L. *Nature* **270**, 66–68 (2011).
2. Johnson, K. E., Leitherer, C., Vacca, W. D. & Conti, P. S. *Astron. J.* **120**, 1273–1288 (2000).
3. Allen, D. A., Wright, A. E. & Goss, W. M. *Mon. Not. R. Astron. Soc.* **177**, 91–97 (1976).
4. Filippenko, A. V. & Ho, L. C. *Astrophys. J.* **588**, L13–L16 (2003).
5. Barth, A. J. *et al. Astrophys. J.* **607**, 90–102 (2004).
6. Greene, J. E. & Ho, L. C. *Astrophys. J.* **670**, 92–104 (2007).

7. Desroches, L.-B. & Ho, L. C. *Astrophys. J.* **690**, 267–278 (2009).
8. Satyapal, S., Vega, D., Dudik, R. P., Abel, N. P. & Heckman, T. *Astrophys. J.* **677**, 926–942 (2008).
9. Alvarez, M. A., Wise, J. H. & Abel, T. *Astrophys. J.* **701**, L133–L137 (2009).
10. van der Marel, R. P. in *Coevolution of Black Holes and Galaxies* (ed. Ho, L. C.) 37–52 (Cambridge Univ. Press, 2004).

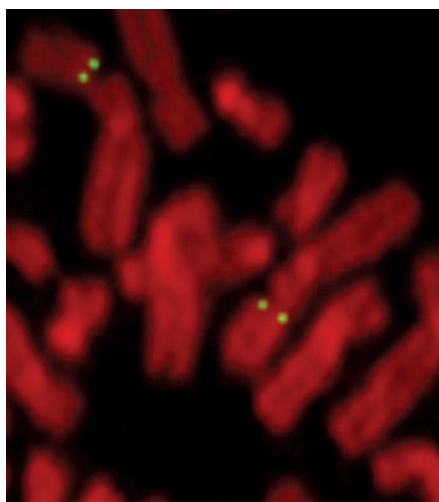
only by convergence of replication forks from the flanking regions (see Fig. 3a of the paper on page 122). The authors propose, therefore, that common chromosome fragile sites correspond to the initiation-poor regions that finish replication last in the given cell type.

To visualize DNA-replication dynamics in the fragile regions, Letessier *et al.* analysed stretched-out DNA strands, and searched for newly synthesized, fluorescently tagged *FRA3B* regions. They could thus examine both replication-fork speed and fork slowing, as well as mapping replication start and stop sites within *FRA3B* in lymphoblasts. The authors report that, within this fragile site, forks did not stall. Instead, the 700-kilobase core region of *FRA3B*, which contains the protein-coding portion of *FHIT* — the region most often mutated by deletion in cancers and precancerous lesions — showed no initiation events. By comparison, roughly ten initiation events occurred in similarly sized, non-fragile regions. These results suggest that a paucity of initiation events contributes to fragility.

A previous genome-wide analysis of replication timing<sup>5</sup> in lymphoblasts and fibroblasts revealed that, in these two cell types, the dynamics of replication (its timing and initiation patterns) are very different. For example, whereas initiation events occurred within the fragile sites in fibroblasts, they were consistently scarce in lymphoblasts. Do such differences in initiation of replication between the two cell types — at least when grown in culture — mean that the fragility of *FRA3B* also differs between them? According to Letessier and co-workers, the answer is indeed yes.

The authors find that fibroblasts do not exhibit the breaks at *FRA3B* that are characteristically seen in lymphoblasts. They also report similar findings for the second most active fragile region known in lymphocytes, *FRA16D*. Together, these data indicate that common chromosome fragile regions are “loci that correspond to the latest initiation-poor regions to complete replication in a given cell type”<sup>1</sup>. How such differential replication programs in cells of different tissue origin are established is unknown, but it cannot depend on DNA sequence because the DNA sequence of the *FRA3B* locus is the same in fibroblasts and lymphoblasts; likewise, the sequence of *FRA16D* is the same in the two cell types.

Common chromosomal fragile sites have previously been assessed in lymphoblasts,



**Figure 1 | Damage-prone DNA.** An image of normal chromosomes in a human lymphoblast. The *FRA3B* region (tagged by a green DNA probe flanking the gaps on both chromosomes 3) is a fragile site prone to breakage. Letessier *et al.*<sup>1</sup> show that one reason for the fragility of this genomic region is a paucity of initiation sites for DNA replication. (Image adapted from ref. 10.)

most active fragile site in a type of human white blood cell called a lymphoblast and which lies within a tumour-suppressor gene known as *FHIT* (Fig. 1). They find that *FRA3B* is a fragile site not because the replication fork is slowed or stalled at this genomic region (locus), but because of the scarcity of replication initiation events there.

Indeed, the study shows that in lymphoblasts — but not in fibroblasts, which originate from connective tissue — initiation events are entirely absent from the central fragile region of *FRA3B*. Consequently, the replication of this large region within *FHIT* can be completed

fibroblasts, some cancer cells of epithelial origin<sup>6,7</sup> and even in cells of different lineages in a single study<sup>8</sup>. What has been lacking, however, is a systematic comparison of the frequency of breaks at known fragile sites in different cell types from a range of organs and tissues. Yet it is known that *FHIT* and *WWOX* — the two genes at the two most fragile loci in lymphoblasts — are also among the sequences most frequently altered by DNA deletion in precancerous and cancerous cells of epithelial origin (that is, cells from internal organs, not lymphoblasts); this is presumably due to exposure of cells within lung, colon and breast to agents that cause replicative stress<sup>6,7</sup>.

It has been argued that damage to genes at fragile sites, and the consequent loss of the genes' expression, contributes to the selective growth of precancerous lesions and cancerous tumours. A counter-argument is that, because of the frequent deletions within the fragile sites, the loss of any associated gene expression is an unselected 'passenger' event in cancers<sup>9</sup> and does not drive the expansive growth of the cancer cells.

A comparison of the frequency of breaks at known fragile sites in a range of cell types, including the epithelial cells of the lung, colon, breast and prostate — where most human cancers originate — could shed light on several questions arising from the current study<sup>1</sup>. For instance, are *FRA3B* and *FRA16D* the most fragile regions in these cells, as they are in lymphoblasts? If fibroblasts show infrequent DNA

damage at these sites, do they contain a different set of fragile sites, or do they show a low frequency of breaks across all chromosomes? Do the frequency and sites of breaks at fragile regions in cultured cells correspond to those in the same cell type within its organ of origin? Finally, if *FRA3B* is not the most fragile site in epithelial cells, does that strengthen the argument that breaks in *FHIT* in precancerous cells and cancers have contributed to progressive growth of the lesions? Undoubtedly, Letessier and colleagues, and indeed other scientists, will be searching for the answers. ■

Kay Huebner is in the Department of Molecular Virology, Immunology and Medical Genetics, Ohio State University Comprehensive Cancer Center, Columbus, Ohio 43210, USA.  
e-mail: kay.huebner@osumc.edu

1. Letessier, A. et al. *Nature* **470**, 120–123 (2011).
2. Durkin, S. G. & Glover, T. W. *Annu. Rev. Genet.* **41**, 169–192 (2007).
3. Le Beau, M. M. et al. *Hum. Mol. Genet.* **7**, 755–761 (1998).
4. Jiang, Y. et al. *Hum. Mol. Genet.* **18**, 4501–4512 (2009).
5. Hansen, R. S. et al. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
6. Tsantoulis, P. K. et al. *Oncogene* **27**, 3256–3264 (2008).
7. Durkin, S. G. et al. *Proc. Natl Acad. Sci. USA* **105**, 246–251 (2008).
8. Turner, B. et al. *Cancer Res.* **62**, 4054–4060 (2002).
9. Bignell, G. R. et al. *Nature* **463**, 893–898 (2010).
10. Zimonjic, D. B. et al. *Cancer Res.* **57**, 1166–1170 (1997).

## THEORETICAL ECOLOGY

# Waltz of the weevil

The aquatic plant *Salvinia molesta* is a widespread pest of waterways in the tropics and subtropics. A study of its control by a weevil in Australian billabongs sets a new standard in ecological time-series analysis. [SEE LETTER P.86](#)

LEWIS STONE

**W**altzing Matilda, the bush ballad that became Australia's unofficial national anthem, relates the exploits of an out-of-control sheep shearer "camped by a billabong" (a small stagnant lake) in the Australian outback. On page 86 of this issue we read about the findings of an inspired group of theoretical ecologists<sup>1</sup> and their models of out-of-control billabongs. Schooler et al.<sup>1</sup> present a mathematical modelling study of an invasive plant species, the noxious weed *Salvinia molesta*, and its erratic large-scale outbreaks in four billabongs over a period of several decades. The authors' skilfully executed modelling is an imaginative combination of nonlinear dynamics, statistical inference and stochastic time-series analysis.

The story of *Salvinia* has become a classic in the biological invasion literature<sup>2</sup>. This aquatic plant from South America is notorious for its rapid, almost uncontrollable growth, and since 1939 has become a pest in regions far from its home range. The weed is able to double its biomass every 3–4 days, generating thick mats of plant material that often cover entire water bodies (Fig. 1). *Salvinia* infestations can have devastating effects on lakes, billabongs and waterways, rendering them unusable for irrigation, as sources of drinking water or for sustaining local fish populations. In the past, entire villages in Papua New Guinea have been abandoned and the inhabitants relocated as a consequence of this out-of-control weed.

*Salvinia* first appeared in Australia's Kakadu National Park in 1983 and, within months, biological control was implemented by the



## 50 Years Ago

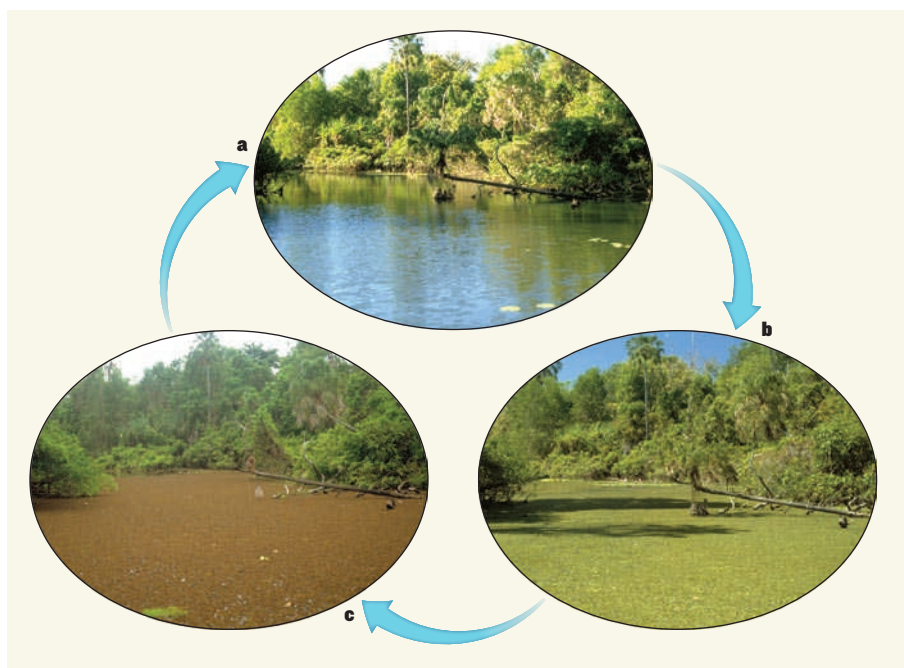
The Committee on Scientific Research in Schools which was established by the Council of the Royal Society in 1957 has issued a report covering the period November 1, 1959–October 31, 1960. It is stated that interest shown by schools in undertaking research has continued to increase, and the Committee is now administering research projects in 56 schools. Some of these schools have more than one research project under way and a total of 66 separate research projects are now being actively pursued compared with 57 last year ... All research projects with which the Committee is dealing are carried out with the specialist advice and assistance of Fellows of the Society and by others who act as advisers. The Committee invites further requests from school-masters and school-mistresses for assistance in undertaking research.

From *Nature* 4 February 1961

## 100 Years Ago

Many attempts have been made to synchronise the phonograph or gramophone with the kinematograph, so as to be able to reproduce simultaneously the sounds of the voice, as in singing and speech, while the movements of the face and the bodily gestures of the singer or speaker are depicted on the screen ... The difficulties, however, appear to have been surmounted by M. Gaumont ... The details of the method are not fully developed, but they are to be made public without delay ... We may soon have in our homes the *chefs-d'oeuvre* of our theatres played by our best actors, and even lectures by famous professors may not be restricted to their class-rooms ... Such reproductions are to be called *phonoscenes*.

From *Nature* 2 February 1911



**Figure 1 | *Salvinia* infestation of the Howard River, near Darwin, Australia.** **a**, The scene before the advent of *Salvinia* in July 1984. **b**, One month later, with *Salvinia* rampant. **c**, February 1985, and the infestation is dying after attack by the weevil *Cyrtobagous salviniae*. Photographs taken by Colin Wilson (at risk of crocodile attack).

introduction of the weevil *Cyrtobagous salviniae*. Unusually, this beetle can feed on *Salvinia* possibly as fast as the weed can grow, causing the latter's dense mats to turn brown through decomposition (Fig. 1). In some cases the weevil can remove 99% of a large-scale *Salvinia* outbreak — which can comprise tens of thousands of tonnes of biomass — within a year.

Schooler *et al.*<sup>1</sup> elaborate on how the annual flooding of the billabongs further complicates the outbreak dynamics. Flooding tends to flush *Salvinia* downstream, allowing the pest to invade other billabongs or to find refuge in terrestrial sites. The ability of *Salvinia* to escape during flooding makes it almost impossible to eradicate the weed, despite the high efficiency of the biological control. Hence outbreaks recur, appearing at erratic and unpredictable intervals as they manage to evade the weevil's stranglehold.

The jumps from periods of control to periods of outbreaks have allowed Schooler *et al.*<sup>1</sup> to draw upon the powerful theoretical framework of alternative stable states, which has proved particularly relevant to those ecosystems in which abrupt changes and catastrophic shifts are intrinsic features<sup>2–4</sup>. The authors were able to formulate the basic structure of a mathematical model that suits the *Salvinia*–weevil system. The nonlinear model has a deterministic 'skeleton' that is driven by environmental stochasticity, and takes into account the observed time-series measurement errors in *Salvinia* and weevil population abundance. But matching the model to the time series proved to be a formidable challenge. The high stochasticity and unpredictability of the billabong system,

combined with the complex dynamics introduced by periodic flooding, mean that, using most standard techniques, it is extremely difficult to interpret which of the different states the system is moving towards at any point in time.

Modern time-series analysis came to the rescue. Theoretical ecologists are familiar with the contributions of one of the authors (Ives) to ecological time-series analysis<sup>4,5</sup>; Ives and his Australian co-authors now present<sup>1</sup> further innovation. More specifically, they make use of the (extended) Kalman filter, a statistical technique for which its inventor, engineer Rudolf Kalman, received the US National Medal of Science in 2008. The filter smooths out the system's stochasticity and, in parallel, provides an estimate of the model's statistical likelihood.

With such an estimate, it became possible to fine-tune the model structure by comparing a suite of different possibilities, while fitting the models to the observed billabong data. This allowed determination of the most reasonable model, and homed in on the best-fitting parameter estimates in a statistically rigorous manner. These methods are now finding exciting biological applications<sup>6</sup>, but in practice their complexity would normally call for the participation of a versatile mathematician. Hence the importance of multidisciplinary cooperation.

With the final model in hand, Schooler *et al.*<sup>1</sup> initiated a theoretical study of the nonlinear dynamics of its deterministic skeleton and investigated the existence of possible alternative stable states, and the manner in which they depend on control parameters. Two states were



identified: a low, *Salvinia*-free state, and a high state of dense *Salvinia* biomass. Under some conditions, these two stable states can coexist and population trajectories may be attracted to either state. Under other conditions there might be only a single state that is attractive. With flooding events and stochastic forcing, the system may be bouncing in a complex way between states, making it almost impossible to ascertain the underlying rules or patterns just by looking at the *Salvinia* or weevil time series.

The model that emerges from Schooler and colleagues' analysis<sup>1</sup> provides a useful tool for understanding the driving forces behind the *Salvinia*-weevil system — and its alternative stable states — that would otherwise be difficult to identify. With that as background,

the authors discuss how the modelling framework helps to suggest practical solutions for biological control. In particular, they argue that it may be possible to take advantage of the system's stochastic fluctuations and its associated erratic jumping between alternative states. In the higher state, when *Salvinia* biomass is at high density, weevil control is least effective. However, augmenting weevil control at those times when the system is attracted to its lower state might possibly trap the system into a stable *Salvinia*-free state. The model could thus help managers to identify the optimal time to apply biological control.

In all, Schooler and colleagues' careful attention to data, and their development and implementation of modelling techniques,

set a new standard in ecological time-series analysis. Their approach promises to have many applications in future studies of noisy biological data sets. ■

**Lewi Stone** is in the Department of Zoology, Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel.  
e-mail: lewi@post.tau.ac.il

1. Schooler, S. S., Salau, B., Julien, M. H. & Ives, A. R. *Nature* **470**, 86–89 (2011).
2. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. *Nature* **413**, 591–596 (2001).
3. Scheffer, M. *et al.* *Nature* **461**, 53–59 (2009).
4. Ives, A. R. & Carpenter, S. R. *Science* **317**, 58–62 (2007).
5. Ives, A. R., Abbott, K. C. & Ziebarth, N. L. *Ecology* **91**, 858–871 (2010).
6. Dowd, M. J. *Mar. Syst.* **68**, 439–456 (2007).

## GLOBAL CHANGE

# Methane and monsoons

**The rising trend in atmospheric concentrations of methane over the past 5,000 years has been attributed to human agency. A modelling study, of a power that has only now become possible, points to another cause. SEE LETTER P.82**

ERIC W. WOLFF

**M**ethane is a potent greenhouse gas, and influences the levels of other atmospheric constituents. The huge increase, of about 150%, to nearly 1,800 parts per billion by volume (p.p.b.v.) in atmospheric concentration over the past two centuries<sup>1</sup> is clearly caused by human activities. However, methane concentration also increased significantly, from about 550 to 700 p.p.b.v., over the previous 5,000 years — the later part of the (interglacial) Holocene epoch that began some 10,000 years ago. There has been intense debate about whether this rise was also anthropogenic or was due to changes in natural sources and sinks. Using models of climate, vegetation and emissions, Singarayer *et al.*<sup>2</sup> show how the increase could have arisen from natural causes (page 82 of this issue).

Detailed ice-core data for methane now cover the past 800,000 years<sup>3</sup>. They show a characteristic pattern over glacial–interglacial cycles, with higher values during interglacials. From the combined use of methane-concentration and isotopic data, it seems that the main cause of the glacial–interglacial rise was almost certainly an increase in the strength of wetland methane sources<sup>4</sup>, perhaps allied to a weakening of the atmospheric sink. Rapid fluctuations,

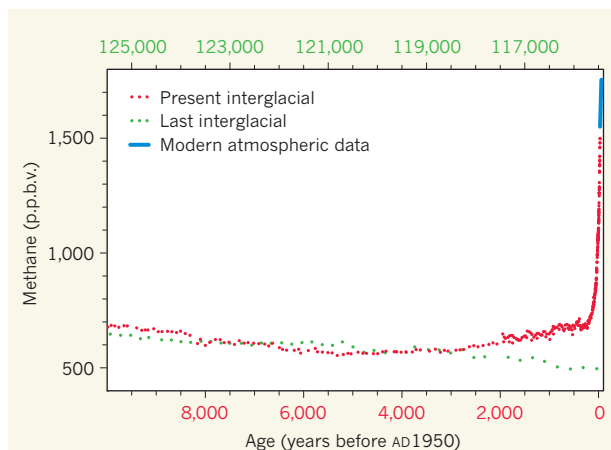
simultaneous with fast, millennial-scale climatic changes in the Northern Hemisphere, are also seen. Finally, the ‘envelope’ of data seems to follow closely the pattern of precession in Earth's orbit, which has a roughly 20,000-year cycle. The apparent reason for this is that tropical wetland emissions of methane respond to the amount of incoming solar

radiation (insolation) in summer at northern low latitudes. Insolation reaches its maximum during the part of the precession cycle when the elliptical orbit of Earth takes the planet closest to the Sun during northern summer. The result is a stronger monsoon in Asia and other regions, with more summer precipitation, and consequently greater wetland areas and methane production by soil-dwelling microorganisms.

However, the increase of the past 5,000 years departed from this pattern, with an increase in atmospheric methane concentration at a time when northern summer insolation was decreasing. In influential papers<sup>5,6</sup>, Ruddiman proposed that the increase was due to human, especially agricultural, activity, which overwhelmed the variations in natural sources, even 5,000 years ago. There has been debate about whether the much smaller human population of that period could really have had such a dominant effect. A strength of the hypothesis has been that the pattern of the past few millennia differed from that of earlier interglacials, which more closely followed the precessional insolation pattern.

Singarayer and colleagues' approach<sup>2</sup> involved an intensive modelling programme. To provide snapshots roughly every 2,000 years over the last glacial cycle, spanning 130,000 years, they ‘forced’ the Hadley Centre's HadCM3 coupled ocean–atmosphere general circulation model (GCM) with the appropriate orbital and ice-sheet configurations, and with greenhouse-gas concentrations, and ran the model to equilibrium. Other simulations were carried out, in which one or more of these forcings was held constant to isolate the causes of change.

The authors then fed the output of each climate simulation through a series of offline models



**Figure 1 | Atmospheric methane concentrations during the present and last interglacials.** Almost all of the data come from ice cores<sup>1,3</sup>. The curves are for the past 10,000 years — the Holocene — and for the equivalent period (125,000–115,000 years ago, in terms of orbital precession) in the last interglacial. The red curve (with modern atmospheric data in blue) shows methane levels during the present interglacial, with a rise commencing 5,000 years ago. The green curve shows the contrasting continual decrease in methane during the last interglacial. Singarayer and colleagues' modelling study<sup>2</sup> can explain the trends in both interglacials in terms of Earth's orbit, except for the past 200 years, when a marked anthropogenic effect has occurred.

# A closely packed system of low-mass, low-density planets transiting Kepler-11

Jack J. Lissauer<sup>1</sup>, Daniel C. Fabrycky<sup>2</sup>, Eric B. Ford<sup>3</sup>, William J. Borucki<sup>1</sup>, Francois Fressin<sup>4</sup>, Geoffrey W. Marcy<sup>5</sup>, Jerome A. Orosz<sup>6</sup>, Jason F. Rowe<sup>7</sup>, Guillermo Torres<sup>4</sup>, William F. Welsh<sup>6</sup>, Natalie M. Batalha<sup>8</sup>, Stephen T. Bryson<sup>1</sup>, Lars A. Buchhave<sup>9</sup>, Douglas A. Caldwell<sup>7</sup>, Joshua A. Carter<sup>4</sup>, David Charbonneau<sup>4</sup>, Jessie L. Christiansen<sup>7</sup>, William D. Cochran<sup>10</sup>, Jean-Michel Desert<sup>4</sup>, Edward W. Dunham<sup>11</sup>, Michael N. Fanelli<sup>12</sup>, Jonathan J. Fortney<sup>2</sup>, Thomas N. Gautier III<sup>13</sup>, John C. Geary<sup>4</sup>, Ronald L. Gilliland<sup>14</sup>, Michael R. Haas<sup>1</sup>, Jennifer R. Hall<sup>15</sup>, Matthew J. Holman<sup>4</sup>, David G. Koch<sup>1</sup>, David W. Latham<sup>4</sup>, Eric Lopez<sup>2</sup>, Sean McCauliff<sup>15</sup>, Neil Miller<sup>2</sup>, Robert C. Morehead<sup>3</sup>, Elisa V. Quintana<sup>7</sup>, Darin Ragozzine<sup>4</sup>, Dimitar Sasselov<sup>4</sup>, Donald R. Short<sup>6</sup> & Jason H. Steffen<sup>16</sup>

**When an extrasolar planet passes in front of (transits) its star, its radius can be measured from the decrease in starlight and its orbital period from the time between transits. Multiple planets transiting the same star reveal much more: period ratios determine stability and dynamics, mutual gravitational interactions reflect planet masses and orbital shapes, and the fraction of transiting planets observed as multiples has implications for the planarity of planetary systems. But few stars have more than one known transiting planet, and none has more than three. Here we report Kepler spacecraft observations of a single Sun-like star, which we call Kepler-11, that reveal six transiting planets, five with orbital periods between 10 and 47 days and a sixth planet with a longer period. The five inner planets are among the smallest for which mass and size have both been measured, and these measurements imply substantial envelopes of light gases. The degree of coplanarity and proximity of the planetary orbits imply energy dissipation near the end of planet formation.**

Kepler is a 0.95-m-aperture space telescope using transit photometry to determine the frequency and characteristics of planets and planetary systems<sup>1–4</sup>. The only fully validated multiple transiting planet system to appear in the literature to date is Kepler-9, with two giant planets<sup>5</sup> orbiting exterior to a planet whose radius is only 1.6 times that of Earth<sup>6</sup>. The Kepler-10 system<sup>7</sup> contains one confirmed planet and an additional unconfirmed planetary candidate. Light curves of five other Kepler target stars, each with two or three (unverified) candidate transiting planets, have also been published<sup>8</sup>. A catalogue of all candidate planets, including targets with multiple candidates, is presented in ref. 35.

We describe below a six-planet system orbiting a star that we name Kepler-11. First, we discuss the spacecraft photometry on which the discovery is based. Second, we summarize the stellar properties, primarily constrained using ground-based spectroscopy. Then we show that slight deviations of transit times from exact periodicity owing to mutual gravitational interactions confirm the planetary nature of the five inner candidates and provide mass estimates. Next, the outer planet candidate is validated by computing an upper bound on the probability that it could result from known classes of astrophysical false positives. We then assess the dynamical properties of the system, including long-term stability, eccentricities and relative inclinations of the planets' orbital planes. We conclude with a discussion of constraints on the compositions of the planets and the clues that the compositions of these planets and their orbital dynamics provide for the structure and formation of planetary systems.

## Kepler photometry

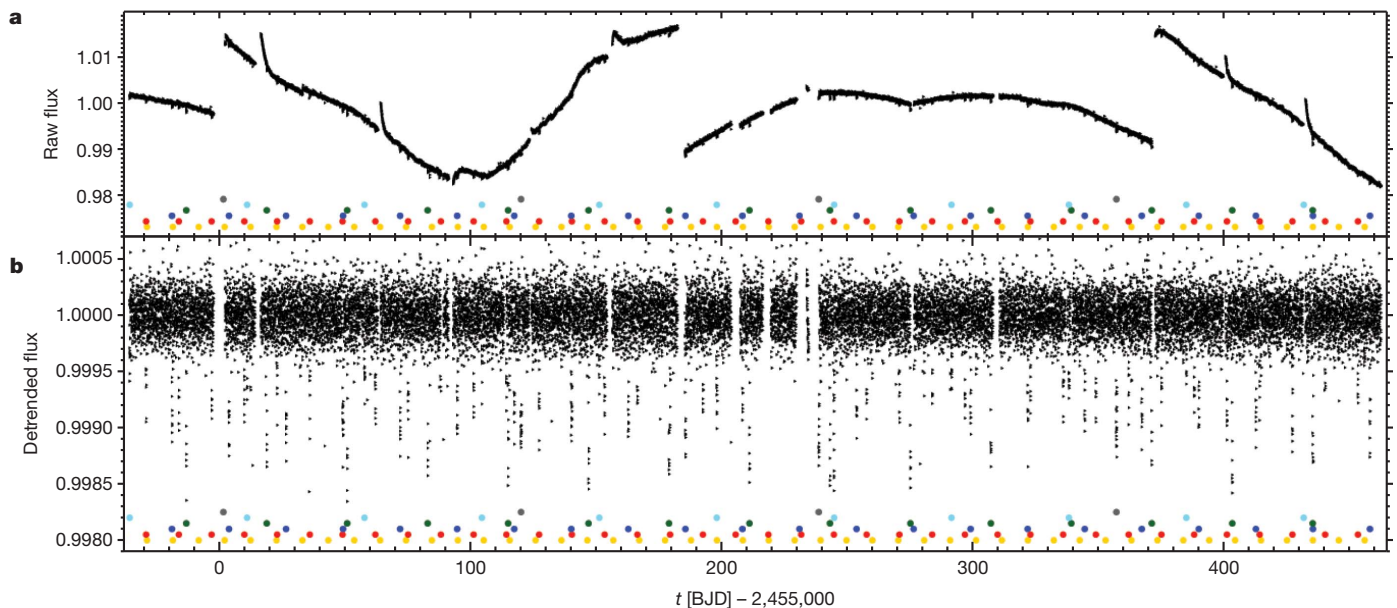
The light curve of the target star Kepler-11 is shown in Fig. 1. After detrending, six sets of periodic dips of depth of roughly one millimagnitude (0.1%) can be seen. When the curves are phased with these six periods, each set of dips (Fig. 2) is consistent with a model<sup>9</sup> of a dark, circular disk masking out light from the same limb-darkened stellar disk; that is, evidence of multiple planets transiting the same star. We denote the planets in order of increasing distance from the star as Kepler-11b, Kepler-11c, Kepler-11d, Kepler-11e, Kepler-11f and Kepler-11g.

Background eclipsing binary stars can mimic the signal of a transiting planet<sup>10</sup>. Kepler returns data for each target as an array of pixels, enabling post-processing on the ground to determine the shift, if any, of the location of the target during the apparent transits. For all six planetary candidates of Kepler-11, these locations are coincident, with  $3\sigma$  uncertainties of 0.7 arcseconds or less for the four largest planets and 1.4 arcseconds for the two smallest planets; see the first section of the Supplementary Information and Supplementary Table 1 for details. This lack of displacement during transit substantially restricts the parameter space available for background eclipsing binary star false positives.

Supplementary Table 2 lists the measured transit depths and durations for each of the planets. The durations of the drops in flux caused by three of the planets are consistent with near-central transits of the same star by planets on circular orbits. Kepler-11e's transits are one-third shorter than expected, implying an inclination to the plane of the sky of  $88.8^\circ$  (orbital eccentricity can also affect transit duration, but

<sup>1</sup>NASA Ames Research Center, Moffett Field, California 94035, USA. <sup>2</sup>UCO/Lick Observatory, University of California, Santa Cruz, California 95064, USA. <sup>3</sup>University of Florida, 211 Bryant Space Science Center, Gainesville, Florida 32611-2055, USA. <sup>4</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. <sup>5</sup>Department of Astronomy, UC Berkeley, Berkeley, California 94720, USA. <sup>6</sup>San Diego State University, 5500 Campanile Drive, San Diego, California 92182, USA. <sup>7</sup>SETI Institute/NASA Ames Research Center, Moffett Field, California 94035, USA.

<sup>8</sup>Department of Physics and Astronomy, San Jose State University, One Washington Square, San Jose, California 95192, USA. <sup>9</sup>Niels Bohr Institute, Copenhagen University, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark. <sup>10</sup>McDonald Observatory, The University of Texas at Austin, Austin, Texas 78712-0259, USA. <sup>11</sup>Lowell Observatory, 1400 W. Mars Hill Road, Flagstaff, Arizona 86001, USA. <sup>12</sup>Bay Area Environmental Research Inst./NASA Ames Research Center, Moffett Field, California 94035, USA. <sup>13</sup>Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA. <sup>14</sup>Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. <sup>15</sup>Orbital Sciences Corporation/NASA Ames Research Center, Moffett Field, California 94035, USA. <sup>16</sup>Fermilab Center for Particle Astrophysics, MS 127, PO Box 500, Batavia, Illinois 60510, USA.



**Figure 1 | Light curves of Kepler-11, raw and detrended.** Kepler-11 is a G dwarf star with Kepler magnitude of 13.7, visual magnitude of 14.2 magnitudes, and celestial coordinates RA = 19 h 48 min 27.62 s, dec. = +41° 54′ 32.9″; alternative designations used in catalogues are KIC 6541920 and KOI-157. Kepler-11 is about 2,000 light-years from Earth. Variations in the brightness of Kepler-11 have been monitored with an effective duty cycle of 91% over the time interval barycentric Julian date (BJD) 2,454,964.511 to 2,455,462.296, with data returned to Earth at a long cadence of 29.426 min. Shown are Kepler photometric data, raw from the spacecraft with each quarter normalized to its median (a) and after detrending with a polynomial filter (b)<sup>31</sup>;  $t$  represents time

the eccentricity needed to explain this duration for a central transit would destabilize the system). The transit durations of planets Kepler-11b and Kepler-11f suggest somewhat non-central transits. In sum, the light curve shapes imply that the system is not perfectly coplanar: Kepler-11g and Kepler-11e are mutually inclined by at least  $\sim 0.6^\circ$ .

### Ground-based spectroscopy

We performed a standard spectroscopic analysis<sup>11,12</sup> of a high-resolution spectrum of Kepler-11 taken at the Keck I telescope. We derive an effective temperature of  $T_{\text{eff}} = 5,680 \pm 100$  K, surface gravity  $g$  of  $\log[g \text{ (cm s}^{-2})] = 4.3 \pm 0.2$ , metallicity of  $[\text{Fe/H}] = 0.0 \pm 0.1$  dex, and an projected stellar equatorial rotation of  $v \sin i = 0.4 \pm 0.5 \text{ km s}^{-1}$ . Combining these measurements with stellar evolutionary tracks<sup>13,14</sup> yields estimates of the star's mass,  $0.95 \pm 0.10 M_\odot$ , and radius,  $1.1 \pm 0.1 R_\odot$ , where the subscript  $\odot$  signifies solar values. Estimates of the stellar density based upon transit observations are consistent with these spectroscopically determined parameters. Therefore, we adopt these stellar values for the rest of the paper, and note that the planet radii scale linearly with the stellar radius. Additional details on these studies are provided in section 2 of the Supplementary Information.

### Transit timing variations

Transits of a single planet on a Keplerian orbit about its star must be strictly periodic. In contrast, the gravitational interactions among planets in a multiple planet system cause planets to speed up and slow down by small amounts, leading to deviations from exact periodicity of transits<sup>15,16</sup>. Such variations are strongest when planetary orbital periods are commensurate or nearly so, which is the case for the giant planets Kepler-9b and Kepler-9c (ref. 5), or when planets orbit close to one another, which is the case for the inner five transiting planets of Kepler-11.

The transit times of all six planets are listed in Supplementary Table 2. Deviations of these times from the best-fitting linear ephemeris (transit timing variations, or TTVs) are plotted in Fig. 3. We modelled

in days since BJD 2,455,000. These data are available from the MAST archive at <http://archive.stsci.edu/kepler/>. Note the difference in vertical scales between the two panels. The six sets of periodic transits are indicated with dots of differing colours. Four photometric data points representing the triply concurrent transit of planets Kepler-11b, Kepler-11d and Kepler-11e at BJD 2,455,435.2 (Supplementary Fig. 12) are not shown because their values lie below the plotted range. Data have also been returned for this target star at a cadence of 58.85 s since BJD 2,455,093.215, but our analysis is based exclusively on the long cadence data.

these deviations with a system of coplanar, gravitationally interacting planets using numerical integrations<sup>5,17</sup> (Supplementary Information). The TTVs for each planet are dominated by the perturbations from its immediate neighbours (Supplementary Fig. 5). The relative periods and phases of each pair of planets, and to a lesser extent the small eccentricities, determine the shapes of the curves in Fig. 3; the mass of each perturber determines the amplitudes. Thus this TTV analysis allows us to estimate the masses of the inner five planets and to place constraints on their eccentricities. We report the main results in Table 1 and detailed fitting statistics in Supplementary Fig. 5 and associated text).

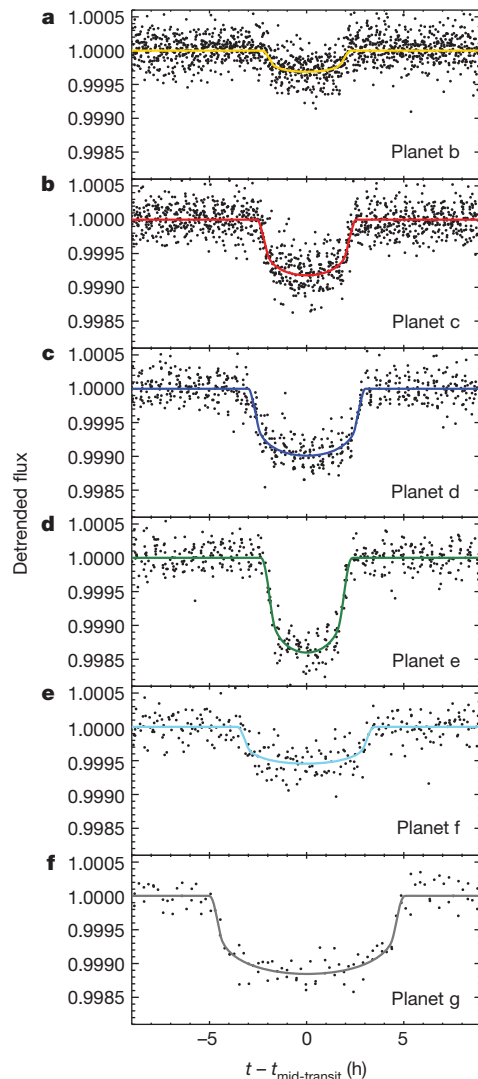
Perturbations of planets Kepler-11d and Kepler-11f by planet Kepler-11e are clearly observed. These variations confirm that all three sets of transits are produced by planets orbiting the same star and yield a  $4\sigma$  detection of the mass of Kepler-11e. Somewhat weaker perturbations are observed in the opposite direction, yielding a  $3\sigma$  detection of the mass of Kepler-11d and a  $2\sigma$  detection of the mass of Kepler-11f.

The inner pair of observed planets, Kepler-11b and Kepler-11c, lie near a 5:4 orbital period resonance and strongly interact with one another. The degree to which they deviate from exact resonance determines the frequency at which their TTVs should occur. Even though the precision of individual transit times is low owing to small transit depths, transit-timing periodograms of both planets show peak power at the expected frequency (Supplementary Fig. 4). The TTVs thus confirm that Kepler-11b and Kepler-11c are planets, confirm that they orbit the same star, and yield  $2\sigma$  determinations of their masses. The outer planet, Kepler-11g, does not strongly interact with the others; it would need to be unexpectedly massive ( $\sim 1 M_{\text{Jupiter}}$ ) to induce a detectable ( $\Delta\chi^2 = 9$ ) signal on the entire set of transit mid-times.

### Validation of planet Kepler-11g

The outer planetary candidate is well-separated from the inner five in orbital period, and its dynamical interactions are not manifested in the



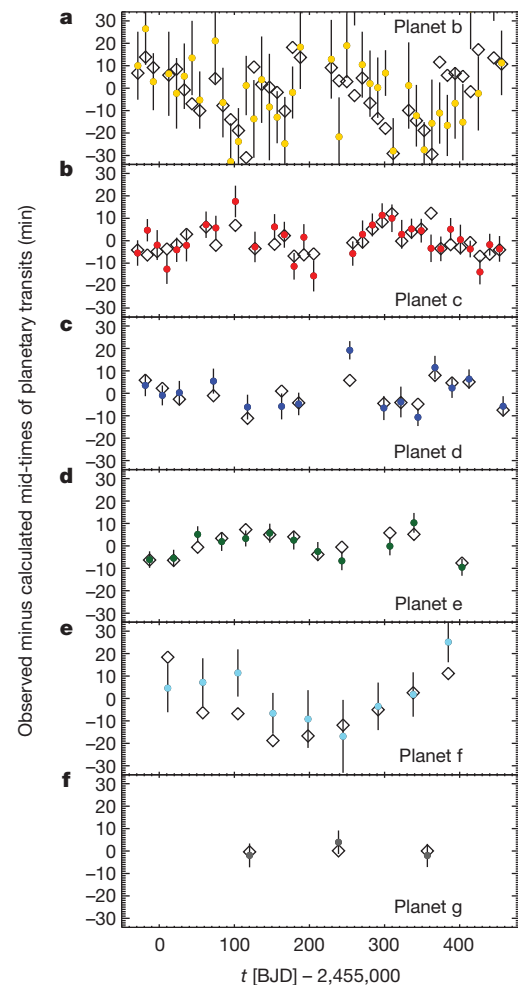


**Figure 2 | Detrended data** of Fig. 1 shown phased at the period of each transit signal and zoomed to an 18-h region around mid-transit.

Overlapping transits are not shown, nor were they used in the model. Each panel has an identical vertical scale, to show the relative depths, and identical horizontal scale, to show the relative durations. The colour of each planet's model light curve matches the colour of the dots marking its transits in Fig. 1.

data presently available. Thus, we only have a weak upper bound on its mass, and unlike the other five candidates, its planetary nature is not confirmed by dynamics. The signal (Fig. 2f) has the characteristics of a transiting planet and is far too large to have a non-negligible chance of being due to noise, but the possibility that it could be an astrophysical false-positive must be addressed. To obtain a Bayesian estimate of the probability that the events seen are due to a sixth planet transiting the star Kepler-11, we must compare estimates of the a priori likelihood of such a planet and of a false positive. This is the same basic methodology as was used to validate planet Kepler-9d (ref. 6).

We begin by using the BLENDER code<sup>6</sup> to explore the wide range of false positives that might mimic the Kepler-11g signal, by modelling the light curve directly in terms of a blend scenario. The overwhelming majority of such configurations are excluded by BLENDER. We then use all other observational constraints to rule out additional blends, and we assess the a priori likelihood of the remaining false positives. Two classes of false positives were considered: (1) the probability of an eclipsing pair of objects that is physically associated with Kepler-11 providing as good a fit to the Kepler data as provided by a planet transiting the primary star was found to be  $0.31 \times 10^{-6}$ ; (2) the probability that a background eclipsing binary or star + planet pair yielding



**Figure 3 | Transit timing variations and dynamical fits.** Observed mid-times of planetary transits (see section 3 of the Supplementary Information for transit-fitting method and Supplementary Table 2 for transit times) minus a calculated linear ephemeris, are plotted as dots with  $1\sigma$  error bars; colours correspond to the planetary transit signals in Figs 1 and 2. The times derived from the 'circular fit' model described in Supplementary Table 4 are given by the open diamonds. Contributions of individual planets to these variations are shown in Supplementary Fig. 6a.

a signal of appropriate period, depth and shape could be present and not have been detected as a result of a centroid shift in the in-transit data, or other constraints from spectroscopy and photometry, was found to be  $0.58 \times 10^{-6}$ . Thus the total a priori probability of a signal mimicking a planetary transit is  $0.89 \times 10^{-6}$ . There is a  $0.5 \times 10^{-3}$  a priori probability of a transiting sixth planet in the mass-period domain. This value was conservatively estimated (not accounting for the coplanarity of the system; the value would increase by an order of magnitude if we were to assume an inclination distribution consistent with seeing transits of the five inner planets) using the observed distribution of extrasolar planets<sup>18,19</sup>. Details of these calculations are presented in section 5 of the Supplementary Information and Supplementary Figs 8–11. Taking the ratio of these probabilities yields a total false alarm probability of  $1.8 \times 10^{-3}$ , which is small enough for us to consider Kepler-11g to be a validated planet.

### Long-term stability and coplanarity

One of the most striking features about the Kepler-11 system is how close the orbits of the planets are to one another. From suites of numerical integrations<sup>20</sup>, dynamical survival of systems with more than three comparably spaced planets for at least  $10^{10}$  orbits has been shown only if the spacing between orbital semi-major axes ( $a_o - a_i$ ) exceeds a critical

**Table 1 | Planet properties**

Planet	Period (days)	Epoch (BJD)	Semi-major axis (AU)	Inclination (°)	Transit duration (h)	Transit depth (millimagnitude)	Radius ( $R_{\oplus}$ )	Mass ( $M_{\oplus}$ )	Density ( $\text{g cm}^{-3}$ )
b	$10.30375 \pm 0.00016$	$2,454,971.5052 \pm 0.0077$	$0.091 \pm 0.003$	$88.5^{+1.0}_{-0.6}$	$4.02 \pm 0.08$	$0.31 \pm 0.01$	$1.97 \pm 0.19$	$4.3^{+2.2}_{-2.0}$	$3.1^{+2.1}_{-1.5}$
c	$13.02502 \pm 0.00008$	$2,454,971.1748 \pm 0.0031$	$0.106 \pm 0.004$	$89.0^{+1.0}_{-0.6}$	$4.62 \pm 0.04$	$0.82 \pm 0.01$	$3.15 \pm 0.30$	$13.5^{+4.8}_{-6.1}$	$2.3^{+1.3}_{-1.1}$
d	$22.68719 \pm 0.00021$	$2,454,981.4550 \pm 0.0044$	$0.159 \pm 0.005$	$89.3^{+0.6}_{-0.4}$	$5.58 \pm 0.06$	$0.80 \pm 0.02$	$3.43 \pm 0.32$	$6.1^{+3.1}_{-1.7}$	$0.9^{+0.5}_{-0.3}$
e	$31.99590 \pm 0.00028$	$2,454,987.1590 \pm 0.0037$	$0.194 \pm 0.007$	$88.8^{+0.2}_{-0.2}$	$4.33 \pm 0.07$	$1.40 \pm 0.02$	$4.52 \pm 0.43$	$8.4^{+2.5}_{-1.9}$	$0.5^{+0.2}_{-0.2}$
f	$46.68876 \pm 0.00074$	$2,454,964.6487 \pm 0.0059$	$0.250 \pm 0.009$	$89.4^{+0.3}_{-0.2}$	$6.54 \pm 0.14$	$0.55 \pm 0.02$	$2.61 \pm 0.25$	$2.3^{+2.2}_{-1.2}$	$0.7^{+0.7}_{-0.4}$
g	$118.37774 \pm 0.00112$	$2,455,120.2901 \pm 0.0022$	$0.462 \pm 0.016$	$89.8^{+0.2}_{-0.2}$	$9.60 \pm 0.13$	$1.15 \pm 0.03$	$3.66 \pm 0.35$	$<300$	–

$R_{\oplus}$ , radius of the Earth;  $M_{\oplus}$ , mass of the Earth. Planetary periods and transit epochs are the best-fitting linear ephemerides. Periods are given as viewed from the barycentre of our Solar System. Because Kepler-11 is moving towards the Sun with a radial velocity of  $57 \text{ km s}^{-1}$  (Supplementary Fig. 1), actual orbital periods in the rest frame of Kepler-11 are a factor of 1.00019 times as long as the values quoted. Uncertainty in epoch is median absolute deviation of transit times from this ephemeris; uncertainty in period is this quantity divided by the number of orbits between the first and last observed transits. Radii are from Supplementary Table 2; uncertainties represent  $1\sigma$  ranges, and are dominated by uncertainties in the radius of the star. The mass estimates are the uncertainty-weighted means of the three dynamical fits (Supplementary Table 4) to TTV observations (Supplementary Table 2) and the quoted ranges cover the union of  $1\sigma$  ranges of these three fits. One of these fits constrains all of the planets to be on circular orbits, the second one allows only planets Kepler-11b and Kepler-11c to have eccentric orbits, and the third solves for the eccentricities of all five planets Kepler-11b to Kepler-11f; see section 4 of the Supplementary Information. Stability considerations may preclude masses near the upper ends of the quoted ranges for the closely spaced inner pair of planets. Inclinations are with respect to the plane of the sky.

number ( $\mathcal{A}_{\text{crit}} \gtrsim 9$ ) of mutual Hill sphere radii ( $(M_i + M_o)/3M_{\star})^{1/3}(a_i + a_o)/2$ , where the subscripts i and o refer to the inner and outer planets, respectively, and  $\star$  refers to the star (here Kepler-11). All of the observed pairs of planets satisfy this criterion, apart from the inner pair, Kepler-11b and Kepler-11c (section 4.1 of the Supplementary Information). These two planets are far enough from one another to be Hill stable in the absence of other bodies<sup>21</sup> (that is, in the three-body problem), and they are distant enough from the other planets that interactions between the subsystems are likely to be weak. Thus stability is possible, although by no means assured. So we integrated several systems that fit the data (given in Supplementary Table 4) for  $2.5 \times 10^8$  years, as detailed in section 4.1 of the Supplementary Information. Weak chaos is evident both in the mean motions and the eccentricities (Supplementary Fig. 7), but the variations are at a low enough level to be consistent with long-term stability.

It is also of interest to determine whether this planetary system truly is as nearly coplanar as the Solar System, or perhaps even more so. Given that the planets all transit the star, they individually must have nearly edge-on orbits. As discussed above and quantified in Supplementary Table 3, the duration of planet Kepler-11e's transit implies an inclination to the plane of the sky of  $88.8^\circ$ , those of the two innermost planets suggest comparable inclinations, whereas those of the three other planets indicate values closer to  $90^\circ$ . But even though each of the planetary orbits is viewed nearly edge-on, they could be rotated around the line of sight and mutually inclined to each other. The more mutually inclined a given pair of planets is, the smaller the probability that multiple planets transit<sup>22,23</sup>. We therefore ran Monte Carlo simulations to assess the probability of a randomly positioned observer viewing transits of all five inner planets assuming that relative planetary inclinations were drawn from a Rayleigh distribution about a randomly selected plane. The results, displayed in Fig. 4 and Supplementary Table 6, suggest a mean mutual inclination of  $1\text{--}2^\circ$ . Details of these calculations are provided in section 6 of the Supplementary Information.

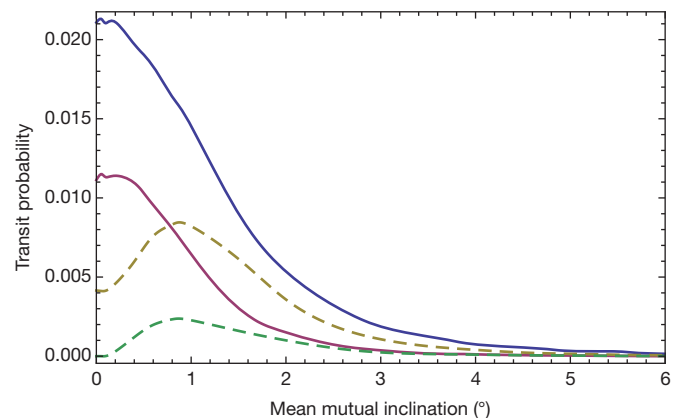
Mutual inclinations around the line of sight give rise to inclination changes, which would manifest themselves as transit duration changes<sup>24</sup>. We notice no such changes. The short baseline, small signal-to-noise ratio and small planet masses render these dynamical constraints weak at the present time for all planets but Kepler-11e. The only planet in the system with an inclination differing significantly from  $90^\circ$  is Kepler-11e, and we find that the duration of its transits does not change by more than 2% over the time span of the light curve. If planet Kepler-11e's orbit were rotated around the line of sight by just  $2^\circ$  compared to all the other components of the system, then with the masses listed in Table 1 the other planets would exert sufficient torque on its orbit to violate this limit.

## Planet compositions and formation

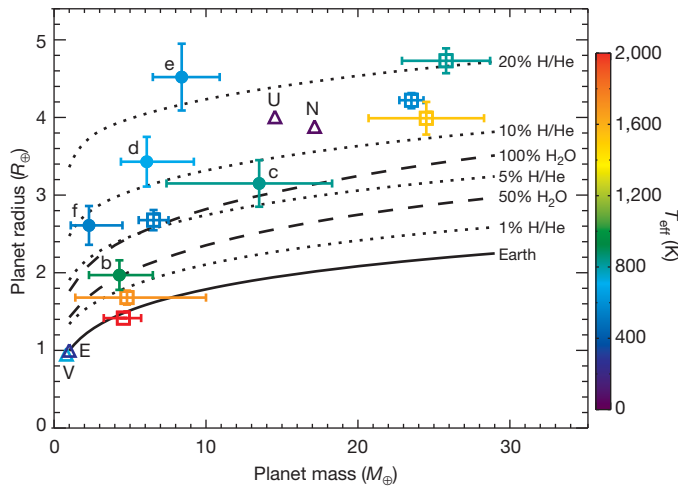
Although the Kepler-11 planetary system is extraordinary, it also tells us much about the ordinary. Measuring both the radii and masses of

small planets is extremely difficult, especially for cooler worlds farther from their star that are not heated above 1,000 K. (Very high temperatures can physically alter planets, producing anomalous properties.) The planetary sizes obtained from transit depths and planetary masses from dynamical interactions together yield insight into planetary composition.

Figure 5 plots radius as a function of mass for the five newly discovered planets the masses of which have been measured. Compared to Earth, each of these planets is large for its mass. Most of the volume of each of the planets Kepler-11c to Kepler-11f is occupied by low-density material. It is often useful to think of three classes of planetary materials, from relatively high to low density: rocks/metals, 'ices' dominated by  $\text{H}_2\text{O}$ ,  $\text{CH}_4$  and  $\text{NH}_3$ , and H/He gas. All of these components could have been accumulated directly from the protoplanetary disk during planet formation. Hydrogen and steam envelopes can also be the product of chemical reactions and out-gassing of rocky planets, but only up to 6% and 20% by mass, respectively<sup>25</sup>. In the Kepler-11



**Figure 4 | Transit probabilities as a function of relative orbital inclinations of planets orbiting Kepler-11.** Results of Monte Carlo simulations to assess the probability of a randomly positioned observer viewing transits of various combinations of observed and hypothesized planets around the star Kepler-11, assuming that relative planetary inclinations were drawn from a Rayleigh distribution about a randomly selected plane. The solid blue curve shows the probability of all five inner planets (Kepler-11b to Kepler-11f) to be seen transiting. The solid pink curve shows the probability of all six observed planets to be seen transiting. The ratio of the orbital period of planet Kepler-11g to that of Kepler-11f is substantially larger than that for any other neighbouring pair of transiting planets in the system. If we hypothesize that a seventh planet orbits between these objects with a period equal to the geometric mean of planets Kepler-11f and Kepler-11g, then the probability of observing transits of any combination totalling six of these seven planets is shown in the dashed golden curve. The dashed green curve shows the probability of the specific observed six planets to be seen transiting. Details of these calculations are provided in section 6 of the Supplementary Information, and numerical results are given in Supplementary Table 6.



**Figure 5 | Mass-radius relationship of small transiting planets, with Solar System planets shown for comparison.** Planets Kepler-11b to Kepler-11f are represented by filled circles with  $1\sigma$  error bars, with their letters written above; values and ranges are as given in Table 1. Other transiting exoplanets in this size range are shown as open squares, representing, in order of ascending radius, Kepler-10b, CoRoT-7b, GJ 1214b, Kepler-4b, GJ 436b and HAT-P-11b. The triangles (labelled V, E, U and N) correspond to Venus, Earth, Neptune and Uranus, respectively. The colours of the points show planetary temperatures (measured for planets in our Solar System, computed mean planet-wide equilibrium temperatures for Bond albedo = 0.2 for the extrasolar planets), with values shown on the colour scale on the right. Using previously implemented planetary structure and evolution models<sup>32,33</sup>, we plot mass-radius curves for 8-Gyr-old planets, assuming  $T_{\text{eff}} = 700$  K. The solid black curve corresponds to models of planets with Earth-like rock-iron composition. The higher dashed curve corresponds to 100%  $\text{H}_2\text{O}$ , using the SESAME 7154  $\text{H}_2\text{O}$  equation of state. All other curves use a water or  $\text{H}_2/\text{He}$  envelope on top of the rock-iron core. The lower dashed curve is 50%  $\text{H}_2\text{O}$  by mass, while the dotted curves are  $\text{H}_2/\text{He}$  envelopes that make up 2%, 6%, 10% and 20% of the total mass. There is significant degeneracy in composition constrained only by mass and radius measurements<sup>34</sup>. Planets Kepler-11d, Kepler-11e and Kepler-11f appear to require a  $\text{H}_2/\text{He}$  envelope, much like Uranus and Neptune, while Kepler-11b and Kepler-11c may have  $\text{H}_2\text{O}$  and/or  $\text{H}_2/\text{He}$  envelopes. We note that multi-component and mixed compositions (not shown above), including rock/iron,  $\text{H}_2\text{O}$ , and  $\text{H}_2/\text{He}$ , are expected and lead to even greater degeneracy in determining composition from mass and radius alone.

system, the largest planets with measured masses, Kepler-11d and Kepler-11e, must contain large volumes of  $\text{H}$ , and low-mass planet Kepler-11f probably does as well. Planets Kepler-11b and Kepler-11c could be rich in ‘ices’ (probably in the fluid state, as in Uranus and Neptune) and/or a  $\text{H}/\text{He}$  mixture. (The error bars on mass and radius for Kepler-11b allow for the possibility of an iron-depleted nearly pure silicate composition, but we view this as highly unlikely on cosmogonic grounds.) In terms of mass, all five of these planets must be primarily composed of elements heavier than helium. Future atmospheric characterization to distinguish between  $\text{H}$ -dominated or steam atmospheres would tell us more about the planets’ bulk composition and atmospheric stability<sup>26</sup>.

Planets Kepler-11b and Kepler-11c have the largest bulk densities and would need the smallest mass fraction of hydrogen to fit their radii. Using an energy-limited escape model<sup>27</sup>, we estimate a hydrogen mass-loss rate of several times  $10^9 \text{ g s}^{-1}$  for each of the five inner planets, leading to the loss of  $\sim 0.1$  Earth masses of hydrogen over 10 Gyr. This is less than a factor of ten below total atmosphere loss for several of the planets. The modelling of hydrogen escape for strongly irradiated exoplanets is not yet well-constrained by observations<sup>28,29</sup>, so larger escape rates are possible. This suggests the scenario that planets Kepler-11b and Kepler-11c had larger  $\text{H}$ -dominated atmospheres in the past and lost these atmospheres during an earlier era when the planets had larger radii, lower bulk density, and a more active primary star, which would all favour higher mass-loss rates.

The comparative planetary science permitted by the planets in Kepler-11 system may allow for advances in understanding these mass-loss processes.

The inner five observed planets of the Kepler-11 planetary system are quite densely packed dynamically, in that significantly closer orbits would not be stable for the billions of years that the star has resided on the main sequence. The eccentricities of these planets are small, and the inclinations very small. The planets are not locked into low-order mean motion resonances.

Kepler-11 is a remarkable planetary system whose architecture and dynamics provide clues to its formation. The significant light-gas component of planets Kepler-11d, Kepler-11e and Kepler-11f imply that at least these three bodies formed before the gaseous component of their protoplanetary disk dispersed, probably taking no longer than a few million years to grow to near their present masses. The small eccentricities and inclinations of all five inner planets imply dissipation during the late stages of the formation/migration process, which means that gas and/or numerous bodies much less massive than the current planets were present. The lack of strong orbital resonances argues against slow, convergent migration of the planets, which would lead to trapping in such configurations, although dissipative forces could have moved the inner pair of planets out from the nearby 5:4 resonance<sup>30</sup>. *In situ* formation would require a massive protoplanetary disk of solids near the star and/or trapping of small solid bodies whose orbits were decaying towards the star as a result of gas drag; it would also require accretion of significant amounts of gas by hot small rocky cores, which has not been demonstrated. (The temperature this close to the growing star would have been too high for ices to have condensed.) The Kepler spacecraft is scheduled to continue to return data on the Kepler-11 planetary system for the remainder of its mission, and the longer temporal baseline afforded by these data will allow for more accurate measurements of the planets and their interactions.

Received 13 December; accepted 20 December 2010.

- Borucki, W. J. *et al.* Kepler planet-detection mission: introduction and first results. *Science* **327**, 977–980 (2010).
- Koch, D. G. *et al.* Kepler mission design, realized photometric performance, and early science. *Astrophys. J.* **713**, L79–L86 (2010).
- Jenkins, J. *et al.* Overview of the Kepler science processing pipeline. *Astrophys. J.* **713**, L87–L91 (2010).
- Caldwell, D. *et al.* Instrument performance in Kepler’s first months. *Astrophys. J.* **713**, L92–L96 (2010).
- Holman, M. J. *et al.* Kepler-9: a system of multiple planets transiting a sun-like, confirmed by timing variations. *Science* **330**, 51–54 (2010).
- Torres, G. *et al.* Modeling Kepler transit light curves as false positives: rejection of blend scenarios for Kepler-9, and validation of Kepler-9d, a super-Earth-size planet in a multiple system. *Astrophys. J.* **727**, 24 (2011).
- Batalha, N. *et al.* Kepler’s first rocky planet: Kepler-10b. *Astrophys. J.* **728**, (in the press).
- Steffen, J. H. *et al.* Five Kepler target stars that show multiple transiting exoplanet candidates. *Astrophys. J.* **725**, 1226–1241 (2010).
- Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
- Brown, T. Expected detection and false alarm rates for transiting Jovian planets. *Astrophys. J.* **593**, L125–L128 (2003).
- Valenti, J. A. & Piskunov, N. Spectroscopy made easy: a new tool for fitting observations with synthetic spectra. *Astron. Astrophys.* **118** (Suppl.), 595–603 (1996).
- Valenti, J. A. & Fischer, D. A. Spectroscopic properties of cool stars (SPOCS). I. 1040 F, G, and K dwarfs from Keck, Lick, and AAT planet search programs. *Astrophys. J.* **159** (Suppl.), 141–166 (2005).
- Girardi, L., Bressan, A., Bertelli, G. & Chiosi, C. Evolutionary tracks and isochrones for low- and intermediate-mass stars: from 0.15 to 7  $M_{\text{sun}}$  and from  $Z=0.0004$  to 0.03. *Astron. Astrophys.* **141** (Suppl.), 371–383 (2000).
- Yi, S. *et al.* Toward better age estimates for stellar populations: The  $Y^2$  isochrones for solar mixture. *Astrophys. J.* **136** (Suppl.), 417–437 (2001).
- Holman, M. J. & Murray, N. W. The use of transit timing to detect terrestrial-mass exoplanets. *Science* **307**, 1288–1291 (2005).
- Agol, E., Steffen, J., Sari, R. & Clarkson, W. On detecting terrestrial planets with timing of giant planet transits. *Mon. Not. R. Astron. Soc.* **359**, 567–579 (2005).
- Fabrycky, D. in *Exoplanets* (ed. Seager, S.) 217–238 (University of Arizona Press, 2010).
- Cumming, A. *et al.* The Keck planet search: detectability and the minimum mass and orbital period distribution of extrasolar planets. *Publ. Astron. Soc. Pac.* **120**, 531–554 (2008).



19. Howard, A. *et al.* The occurrence and mass distribution of close-in super-Earths, Neptunes, and Jupiters. *Science* **330**, 653–655 (2010).
20. Smith, A. W. & Lissauer, J. J. Orbital stability of systems of closely-spaced planets. *Icarus* **201**, 381–394 (2009).
21. Gladman, B. Dynamics of systems of two close planets. *Icarus* **106**, 247–263 (1993).
22. Koch, D. & Borucki, W. A search for earth-sized planets in habitable zones using photometry. *First Int. Conf. Circumstellar Habitable Zones* 229 (Travis House Publishing, 1996).
23. Ragozzine, D. & Holman, M. J. The value of systems with multiple transiting planets. *Astrophys. J.* (submitted); preprint at (<http://arxiv.org/abs/1006.3727>) (2010).
24. Miralda-Escude, J. Orbital perturbations of transiting planets: a possible method to measure stellar quadrupoles and to detect earth-mass planets. *Astrophys. J.* **564**, 1019–1023 (2002).
25. Elkins-Tanton, L. T. & Seager, S. Ranges of atmospheric mass and composition of super-Earth exoplanets. *Astrophys. J.* **685**, 1237–1246 (2008).
26. Miller-Ricci, E., Seager, S. & Sasselov, D. The atmospheric signatures of super-Earths: how to distinguish between hydrogen-rich and hydrogen-poor atmospheres. *Astrophys. J.* **690**, 1056–1067 (2009).
27. Lecavelier des Etangs, A. A diagram to determine the evaporation status of extrasolar planets. *Astron. Astrophys.* **461**, 1185–1193 (2007).
28. Vidal-Madjar, A. *et al.* An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* **422**, 143–146 (2003).
29. Lecavelier des Etangs, A. *et al.* Evaporation of the planet HD 189733 observed in H I Lyman- $\alpha$ . *Astron. Astrophys.* **514**, A72 (2010).
30. Papaloizou, J. C. B. & Terquem, C. On the dynamics of multiple systems of hot super-Earths and Neptunes: tidal circularization, resonance and the HD 40307 system. *Mon. Not. R. Astron. Soc.* **405**, 573–592 (2010).
31. Rowe, J. F. *et al.* Kepler observations of transiting hot compact objects. *Astrophys. J.* **713**, L150–L154 (2010).
32. Miller, N., Fortney, J. J. & Jackson, B. Inflating and deflating hot Jupiters: coupled tidal and thermal evolution of known transiting planets. *Astrophys. J.* **702**, 1413–1427 (2009).
33. Nettelmann, N., Fortney, J. J., Kramm, U. & Redmer, R. Thermal evolution and structure models of the transiting super-Earth GJ 1214b. *Astrophys. J.* (in the press); preprint at (<http://arxiv.org/abs/1010.0277>) (2010).
34. Rogers, L. A. & Seager, S. A framework for quantifying the degeneracies of exoplanet interior compositions. *Astrophys. J.* **712**, 974–991 (2010).
35. Borucki, W. J. *et al.* Characterization of planetary candidates observed by Kepler. *Astrophys. J.* (submitted).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Kepler was competitively selected as the tenth Discovery mission. Funding for this mission is provided by NASA's Science Mission Directorate. We thank the many people who gave so generously of their time to make the Kepler mission a success. A. Dobrovolskis, T. J. Lee and D. Queloz provided constructive comments on the manuscript. D.C.F. and J.A.C. acknowledge NASA support through Hubble Fellowship grants HF-51272.01-A and HF-51267.01-A, respectively, awarded by STScI, operated by AURA under contract NAS 5-26555.

**Author Contributions** J.J.L. led the research effort to confirm and validate candidates as planets, assisted in the dynamical study, developed most of the interpretation and wrote much of the manuscript. D.C.F. performed dynamical analysis on transit times and

derived planetary masses, derived dynamical constraints on mutual inclinations, performed long-term stability calculations, and wrote much of the Supplementary Information. E.B.F. measured transit times, including special processing for Q6 data, checked for transit duration variations, contributed to the interpretation, and supervised transit probability and relative inclination analysis. The following seven authors contributed equally: W.J.B. developed photometers, observational techniques, and analysis techniques that proved Kepler could succeed, participated in the design, development, testing and commissioning of the Kepler mission and in the evaluation of the candidates that led to the discovery of this system. F.F. modelled Kepler transit light curves as false positives leading to the rejection of blend scenarios for hierarchical triple and background configurations. G.W.M. obtained and reduced spectra that yielded the properties of the star. J.A.O. measured planet radii and impact parameters. J.F.R. performed transit searches to identify candidates, multi-candidate light curve modelling to determine stellar and planetary parameters, and transit timing measurements. G.T. modelled Kepler transit light curves as false positives, leading to the rejection of blend scenarios for hierarchical triple and background configurations. W.F.W. measured transit times and O-C curves and used Monte Carlo simulations to determine robust uncertainties. The remaining authors listed below contributed equally: N.M.B. directed target selection, KOI inspection, tracking, and vetting. S.T.B. supported centroid and light curve analysis and participated in validation discussions. L.A.B. took and analysed the first reconnaissance spectrum of the target star. D.A.C. worked on the definition and development of the Science Operations Center analysis pipeline. J.A.C. assisted in the determination of transit times and durations from the Kepler photometry. D.C. provided advice on blender analysis. J.L.C. supported the science operations to collect the Kepler data. W.D.C. obtained, reduced and analysed reconnaissance spectroscopy. J.-M.D. participated in blend studies. E.W.D. provided optical, electronic and systems support for flight segments, commissioning work, and discussions regarding validation of small planets. M.N.F. reviewed light curves and centroid time series and participated in verification and validation of the science pipeline. J.J.F. modelled the interior structure and mass-radius relationships of the planets and wrote the text on interpreting planetary structure. T.N.G. performed follow-up observation support and commissioning work. J.C.G. worked on the design of the Kepler focal plane and associated charge-coupled device (CCD) imagers and electronics. R.L.G. performed difference-image-based centroid analysis as means of discriminating against background eclipsing binary stars. M.R.H. led the team responsible for the scientific commissioning and operation of the instrument, and processing the data to produce light curves. J.R.H. developed operations procedures and processed Kepler data to produce light curves. M.J.H. developed the trending algorithm and helped in assembling and writing up the results. D.G.K. designed and developed major portions of the Kepler mission. D.W.L. led reconnaissance spectroscopy, stellar classification and the preparation of the Kepler Input Catalog. E.L. modelled the interior structure and mass-radius relations of the planets. S.McC. wrote software to manage and archive pixel and flux time series data. N.M. modelled the interior structure and mass-radius relations of the planets. R.C.M. performed transit probability versus relative inclination analysis. E.V.Q. wrote software to calibrate the pixel data to generate the flux time series. D.R. conducted data analysis and interpretation. D.S. performed calculations using stellar evolution models to determine stellar parameters. D.R.S. developed and used code to measure transit times. J.H.S. worked on validation of analysis methods and composition of text.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.J.L. ([jack.lissauer@nasa.gov](mailto:jack.lissauer@nasa.gov)) or D.C.F. ([fabrycky@ucolick.org](mailto:fabrycky@ucolick.org)).

# Mapping copy number variation by population-scale genome sequencing

Ryan E. Mills<sup>1\*</sup>, Klaudia Walter<sup>2\*</sup>, Chip Stewart<sup>3\*</sup>, Robert E. Handsaker<sup>4\*</sup>, Ken Chen<sup>5\*</sup>, Can Alkan<sup>6,7\*</sup>, Alexej Abyzov<sup>8\*</sup>, Seungtae Chris Yoon<sup>9\*</sup>, Kai Ye<sup>10\*</sup>, R. Keira Cheetham<sup>11</sup>, Asif Chinwalla<sup>5</sup>, Donald F. Conrad<sup>2</sup>, Yutao Fu<sup>12</sup>, Fabian Grubert<sup>13</sup>, Iman Hajirasouliha<sup>14</sup>, Fereydoon Hormozdizadeh<sup>14</sup>, Lilia M. Iakoucheva<sup>15</sup>, Zamin Iqbal<sup>16</sup>, Shuli Kang<sup>15</sup>, Jeffrey M. Kidd<sup>6</sup>, Miriam K. Konkel<sup>17</sup>, Joshua Korn<sup>4</sup>, Ekta Khurana<sup>8,18</sup>, Deniz Kural<sup>3</sup>, Hugo Y. K. Lam<sup>13</sup>, Jing Leng<sup>8</sup>, Ruiqiang Li<sup>19</sup>, Yingrui Li<sup>19</sup>, Chang-Yun Lin<sup>20</sup>, Ruibang Luo<sup>19</sup>, Xinmeng Jasmine Mu<sup>8</sup>, James Nemesh<sup>4</sup>, Heather E. Peckham<sup>12</sup>, Tobias Rausch<sup>21</sup>, Aylwyn Scally<sup>2</sup>, Xinghua Shi<sup>1</sup>, Michael P. Stromberg<sup>3</sup>, Adrian M. Stütz<sup>21</sup>, Alexander Ekehart Urban<sup>13,27</sup>, Jerilyn A. Walker<sup>17</sup>, Jiantao Wu<sup>3</sup>, Yujun Zhang<sup>2</sup>, Zhengdong D. Zhang<sup>8</sup>, Mark A. Batzer<sup>17</sup>, Li Ding<sup>5,22</sup>, Gabor T. Marth<sup>3</sup>, Gil McVean<sup>23</sup>, Jonathan Sebat<sup>15</sup>, Michael Snyder<sup>13</sup>, Jun Wang<sup>19,24</sup>, Kenny Ye<sup>20</sup>, Evan E. Eichler<sup>6,7</sup>, Mark B. Gerstein<sup>8,18,25</sup>, Matthew E. Hurles<sup>2</sup>, Charles Lee<sup>1</sup>, Steven A. McCarroll<sup>4,26</sup>, Jan O. Korbel<sup>21</sup> & 1000 Genomes Project†

**Genomic structural variants (SVs) are abundant in humans, differing from other forms of variation in extent, origin and functional impact. Despite progress in SV characterization, the nucleotide resolution architecture of most SVs remains unknown. We constructed a map of unbalanced SVs (that is, copy number variants) based on whole genome DNA sequencing data from 185 human genomes, integrating evidence from complementary SV discovery approaches with extensive experimental validations. Our map encompassed 22,025 deletions and 6,000 additional SVs, including insertions and tandem duplications. Most SVs (53%) were mapped to nucleotide resolution, which facilitated analysing their origin and functional impact. We examined numerous whole and partial gene deletions with a genotyping approach and observed a depletion of gene disruptions amongst high frequency deletions. Furthermore, we observed differences in the size spectra of SVs originating from distinct formation mechanisms, and constructed a map of SV hotspots formed by common mechanisms. Our analytical framework and SV map serves as a resource for sequencing-based association studies.**

## Introduction

Unbalanced structural variants (SVs), or copy number variants (CNVs), involving large-scale deletions, duplications and insertions form one of the least well studied classes of genetic variation. The fraction of the genome affected by SVs is comparatively larger than that accounted for by single nucleotide polymorphisms<sup>1</sup> (SNPs), implying significant consequences of SVs on phenotypic variation. SVs have already been associated with diverse diseases, including autism<sup>2,3</sup>, schizophrenia<sup>4,5</sup> and Crohn's disease<sup>6,7</sup>. Furthermore, locus-specific studies suggest that diverse mechanisms may form SVs *de novo*, with some mechanisms involving complex rearrangements resulting in multiple chromosomal breakpoints<sup>8,9</sup>.

Initial microarray-based SV surveys focused on large gains and losses<sup>10–12</sup>, with recent advances in array technology widening the accessible size spectrum towards smaller SVs<sup>1,13</sup>. Microarray-based surveys commonly mapped SVs to approximate genomic locations. However, a detailed SV characterization, including analyses of SV

origin and impact, requires knowledge of precise SV sequences. Advances in sequencing technology have enabled applying sequence-based approaches for mapping SVs at a fine scale<sup>14–21</sup>. These approaches include: (1) paired-end mapping (or read pair 'RP' analysis) based on sequencing and analysis of abnormally mapping pairs of clone ends<sup>14,22–24</sup> or high-throughput sequencing fragments<sup>15,17,18</sup>; (2) read-depth ('RD') analysis, which detects SVs by analysing the read depth-of-coverage<sup>16,21,25–27</sup>; (3) split-read ('SR') analysis, which evaluates gapped sequence alignments for SV detection<sup>28,29</sup>; and (4) sequence assembly ('AS'), which enables the fine-scale discovery of SVs, including novel (non-reference) sequence insertions<sup>30–32</sup>. Sequence-based SV discovery approaches have previously been applied to a limited (<20) number of genomes, leaving the fine-scale architecture of most common SVs unknown.

Sequence data generated by the 1000 Genomes Project (1000GP) provide an unprecedented opportunity to generate a comprehensive SV map. The 1000GP recently generated 4.1 terabases of raw sequence

<sup>1</sup>Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. <sup>2</sup>The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK. <sup>3</sup>Department of Biology, Boston College, Boston, Massachusetts, USA. <sup>4</sup>Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>5</sup>The Genome Center at Washington University, St. Louis, Missouri, USA. <sup>6</sup>Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA. <sup>7</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA. <sup>8</sup>Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, USA. <sup>9</sup>Seaver Autism Center and Department of Psychiatry, Mount Sinai School of Medicine, New York, New York, USA. <sup>10</sup>Departments of Molecular Epidemiology, Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands. <sup>11</sup>Illumina Cambridge Ltd, Chesterford Research Park, Little Chesterford, Saffron Walden CB10 1XL, UK. <sup>12</sup>Life Technologies, Beverly, Massachusetts, USA. <sup>13</sup>Department of Genetics, Stanford University, Stanford, California, USA. <sup>14</sup>School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada. <sup>15</sup>Department of Psychiatry, Department of Cellular and Molecular Medicine, Institute for Genomic Medicine, University of California, San Diego, La Jolla, California, USA. <sup>16</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. <sup>17</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA. <sup>18</sup>Molecular Biophysics and Biochemistry Department, Yale University, New Haven, Connecticut, USA. <sup>19</sup>BGI-Shenzhen, Shenzhen 518083, China. <sup>20</sup>Albert Einstein College of Medicine, Bronx, New York, USA. <sup>21</sup>Genome Biology Research Unit, European Molecular Biology Laboratory, Heidelberg, Germany. <sup>22</sup>Department of Genetics, Washington University, St. Louis, Missouri, USA. <sup>23</sup>Department of Statistics, University of Oxford, OX3 7BN, UK. <sup>24</sup>Department of Biology, University of Copenhagen, Copenhagen, Denmark. <sup>25</sup>Department of Computer Science, Yale University, New Haven, Connecticut, USA. <sup>26</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. <sup>27</sup>Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA.

\*These authors contributed equally to this work.  
†Lists of participants and affiliations are shown in Supplementary Information.

in two pilot projects targeting whole human genomes<sup>33</sup> (Supplementary Table 1). These studies comprise a population-scale project, termed 'low-coverage project', in which 179 unrelated individuals were sequenced with an average coverage of 3.6×, including 59 Yoruba individuals from Nigeria (YRI), 60 individuals of European ancestry from Utah (CEU), 30 of Han ancestry from Beijing (CHB), and 30 of Japanese ancestry from Tokyo (JPT; the latter two were jointly analysed as JPT+CHB). In addition, a high-coverage project, termed the 'trio project', was carried out, with individuals of a CEU and a YRI parent-offspring trio sequenced to 42× coverage on average.

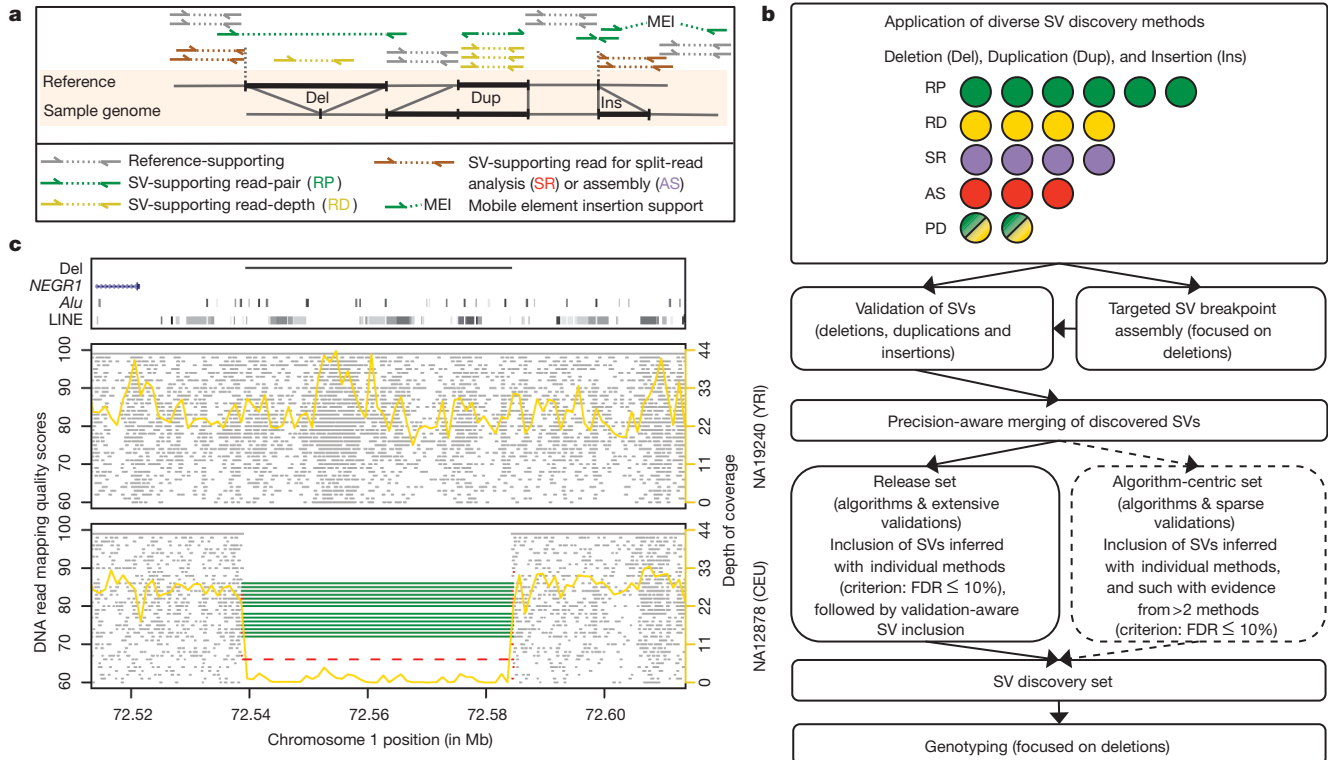
We report here the results of analyses undertaken by the Structural Variation Analysis Group of the 1000GP. The group's objectives were to discover, assemble, genotype and validate SVs of 50 base pairs (bp) and larger in size, and to assess and compare different sequence-based SV detection approaches. The focus of the group was initially on deletions, a variant class often associated with disease<sup>9</sup>, for which rich control data sets and diverse ascertainment approaches exist<sup>1,13,22,28</sup>. Less focus was placed on insertions and duplications<sup>34</sup> and none on balanced SV forms (such as inversions). Specifically, we applied nineteen methods to generate an SV discovery set. We further generated reference genotypes for most deletions, assessed the SVs' functional impact and stratified SV formation mechanism with respect to variant size and genomic context.

### Assessment of SV discovery methods

We incorporated the SV discovery methods into a pipeline (Fig. 1a, b), with the goal of ascertaining different SV types and assessing each method for its ability to discover SVs. The methods detected SVs by analysing RD, RP, SR and AS features, or by combining RP and RD features (abbreviated as 'PD'). Altogether we generated 36 SV call-sets by applying the methods on trio and low-coverage whole genome

sequence data, and by identifying SVs as genomic differences relative to a human reference, corresponding to the reference genome, or to a set of individuals (that is, population reference; Supplementary Table 2). We initially identified SVs as deletions, tandem duplications, novel sequence insertions and mobile element insertions (MEIs) relative to the human reference. Subsequent comparative analyses involving primate genomes enabled us to classify SVs as deletions, duplications, or insertions relative to inferred ancestral genomic loci, reflecting mechanisms of SV formation (see below). DNA reads analysed by SV discovery methods were initially mapped to the human reference genome using a variety of alignment algorithms. Most of these algorithms mapped each read to a single genomic position, although one algorithm (mrFAST<sup>16</sup>) also considered alternative mapping positions for reads aligning to repetitive regions (see Supplementary Tables 2–4 for method-specific parameters and full SV call-sets). We filtered each call-set by excluding SVs <50 bp, which are reported elsewhere<sup>33</sup>. Many SVs showed support from distinct SV discovery methods, as exemplified by a common deletion, previously associated with body-mass index<sup>35</sup> (BMI), that we identified with RP, RD and SR methods (Fig. 1c). Nonetheless, we observed notable differences between methods (Fig. 2a–c) in terms of genomic regions ascertained (Supplementary Fig. 1), accessible SV size-range (Fig. 2a), and breakpoint precision (Fig. 2c, Supplementary Fig. 2).

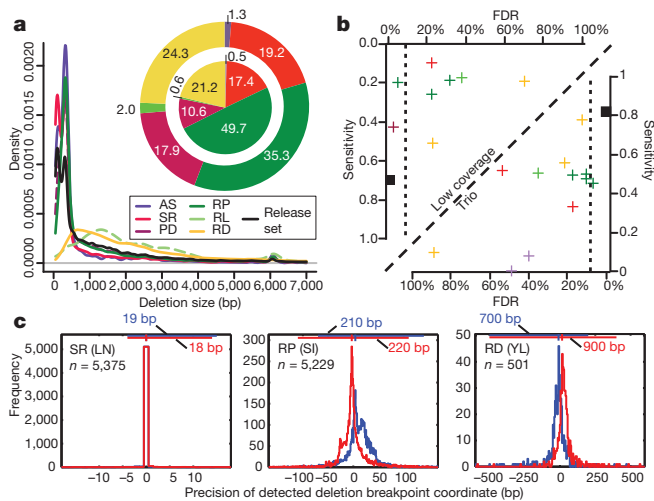
To estimate call-set specificity, we carried out extensive validations (Methods), including PCRs for over 3,000 candidate loci and microarray data analyses for 50,000 candidate loci (Supplementary Tables 3, 4 and Supplementary Fig. 3). We combined PCR and array-based analysis results to estimate false discovery rates (FDRs), and found that eight call-sets (three deletion, one tandem duplication and four insertion call-sets) met the pre-specified specificity threshold<sup>33</sup> (FDR ≤ 10%), whereas the other call-sets yielded lower specificity (FDRs of 13–89%).



**Figure 1 | SV discovery and genotyping in population scale sequence data.** **a**, Schematic depicting the different modes (that is, approaches) of sequence-based SV detection we used. The RP approach assesses the orientation and spacing of the mapped reads of paired-end sequences<sup>14,15</sup> (reads are denoted by arrows); the RD approach evaluates the read depth-of-coverage<sup>25,26</sup>; the SR approach maps the boundaries (breakpoints) of SVs by sequence alignment<sup>28,29</sup>; the AS approach assembles SVs<sup>30–32</sup>. **b**, Integrated pipeline for SV discovery,

validation and genotyping. Coloured circles represent individual SV discovery methods (listed in Supplementary Table 2), with modes indicated by a colour scheme: green, RP; yellow, RD; purple, SR; red, AS; green and yellow, methods evaluating RP and RD (abbreviated as 'PD'). **c**, Example of a deletion, previously associated with BMI<sup>35</sup>, identified independently with RP (green), RD (yellow) and SR (red) methods. Grey dots indicate position and mapping quality for individual sequence reads. Targeted assembly confirmed the breakpoints detected by SR.





**Figure 2 | Comparative assessment of deletion discovery methods.**

**a**, Deletion size-range ascertained by different modes of SV discovery. Three groups are visible, with AS and SR, PD and RP, as well as RD and 'RL' (RP analysis involving relatively long range ( $\geq 1$  kb) insert size libraries, resulting in a different deletion detection size range compared to the predominantly used  $<500$  bp insert size libraries), respectively, ascertaining similar size-ranges. Pie charts display the contribution (%) of different SV discovery modes to the release set. Outer pie is based on the number of SV calls; inner pie is based on the total number of variable nucleotides. Of note, not all approaches were applied across all individuals (see Supplementary Table 2). **b**, Sensitivity and FDR estimates for individual deletion discovery methods based on gold standard sets for individuals sequenced at high (NA12878) and low-coverage (NA12156), respectively. All depicted estimates are summarized in Supplementary Tables 3, 4 and 6. Vertical dotted lines correspond to the specificity threshold (FDR  $\leq 10\%$ ). **c**, Breakpoint mapping resolution of three deletion discovery methods (the respective method names are in Supplementary Table 2). The blue and red histograms are the breakpoint residuals for predicted deletion start and end coordinates, respectively, relative to assembled coordinates (here assessed in low-coverage data). The horizontal lines at the top of each plot mark the 98% confidence intervals (labelled for each panel), with vertical notches indicating the positions of the most probable breakpoint (the distribution mode).

We assessed the sensitivity of deletion discovery methods further by collating data from four earlier surveys<sup>1,13,22,28</sup> into a gold standard (Methods, Supplementary Tables 5, 6 and Supplementary Fig. 4a), and specifically assessing the detection sensitivity for an individual sequenced at high-coverage (NA12878) as well as for an individual sequenced at low-coverage (NA12156). Unsurprisingly, given the typical trade-off between sensitivity and specificity, in the trios the highest sensitivities were achieved by RD and RP methods with FDR  $> 10\%$  (Fig. 2b). By comparison, in the low-coverage data, the individual method with the greatest accuracy (FDR = 3.7%) was the second most sensitive based on our gold standard (Fig. 2b), and the most sensitive when expanding the gold standard to a larger set of individuals (Supplementary Fig. 4b). This method, Genome STRiP (to be described elsewhere; Handsaker, R. E., Korn, J. M., Nemesh, J. and McCarroll, S. A., unpublished results), integrated both RP and RD features (PD), implying that considering different evidence types can improve SV discovery.

### Construction of our SV discovery set

To construct our SV discovery set ('release set'), we joined calls from different discovery methods corresponding to the same SV with a

merging approach that was aware of each call-set's precision in SV breakpoint detection (Supplementary Fig. 5 and Methods). Most SVs in the release set (61%) were contributed by individual methods meeting the pre-defined specificity threshold (FDR  $\leq 10\%$ ). The remaining 39% of calls were contributed by lower specificity methods following experimental validation. Altogether, the release set comprised 22,025 deletions, 501 tandem duplications, 5,371 MEIs and 128 non-reference insertions (Table 1, Supplementary Table 7). With our gold standard we estimated an overall sensitivity of deletion discovery of 82% in the trios, and 69% in low-coverage sequence (Fig. 2b) using a 1-bp overlap criterion. When instead applying a stringent 50% reciprocal overlap criterion for sensitivity assessment (which required SV sizes inferred on different experimental platforms to be in close agreement), our sensitivity estimates decreased by 12% and 18%, respectively, in trio and low-coverage sequence (Supplementary Table 8). We examined further an alternative SV discovery approach that involved the pairwise integration of deletion discovery methods, and tested its ability to discover SVs without relying on the inclusion of lower specificity calls following experimental validation (this approach resulted in the generation of the 'algorithm-centric set'; Fig. 1b). Whereas this alternative approach resulted in an increased number (by  $\sim 13\%$ ) of high-specificity (FDR  $< 10\%$ ) calls compared to the release set (Supplementary Text), overall it resulted in fewer SV calls owing to its decreased sensitivity at the lower ( $<200$  bp) SV size range. In the following analyses we thus focused on the release set.

### Extent and impact of our SV discovery set

We next assessed the extent and impact of our SV discovery (release) set. The median SV size was 729 bp (mean = 8 kilobases), approximately four times smaller than in a recent tiling CGH-based study<sup>1</sup>, reflecting the high resolution of DNA-sequence-based SV discovery. We also compared our set to a recent survey of SVs in an individual genome<sup>36</sup> based on capillary sequencing and array-based analyses<sup>24</sup>, and observed a similar size distribution for deletions, but differences in the size distributions of other SV classes, reflecting underlying differences in SV ascertainment (Supplementary Fig. 6). By comparing our SVs to databases of structural variation and to additional personal genome data sets, we classified 15,556 SVs in our set as novel, with an enrichment of low frequency SVs and small SVs amongst the novel variants (Methods and Supplementary Text).

A major advantage of sequence-based SV discovery is the nucleotide resolution mapping of SVs. We initially mapped the breakpoints of 7,066 deletions and 3,299 MEIs using SR and AS features. Using the TIGRA-targeted assembly approach (Chen, L. *et al.*, unpublished results) we further identified the breakpoints of an additional 4,188 deletions and 160 tandem duplications, initially discovered by RD, RP and PD methods (Methods, Supplementary Tables 3, 4). Altogether, we mapped  $\sim 15,000$  SVs at nucleotide resolution, 48% of which were novel. Few deletion loci (4.4%) displayed different SV breakpoints in different samples, which is explainable by rare TIGRA misassemblies, or alternatively, by recurrently formed, multi-allelic SVs (Supplementary Text). TIGRA further enabled us to validate an additional 7,359 SVs by identifying the SVs' breakpoints (Methods), and to evaluate the mapping precision of SV discovery methods (Fig. 2c, Supplementary Fig. 2).

We assessed the putative functional impact of SVs in our set further by relating them to genomic annotation. Many SVs (1,775) affected coding sequences, resulting in full gene overlaps or exon disruptions (Table 2), many of which led to out-of-frame exons (Supplementary

**Table 1 | Summary of discovered structural variation**

	Deletions	Tandem duplications	Mobile element insertions	Novel sequence insertions	Total
Individual call-sets $< 10\%$ FDR	11,215	501	5,371	—	17,087
Validated experimentally*	10,810	—	—	128	10,938
Release set	22,025	501	5,371	128	28,025

\*Only tabulates validated calls which were not already present in individual call-sets with less than 10% FDR.

**Table 2 | Functional impact of our fine resolution SV set**

SV class	Gene overlap				Total gene overlap	Total intergenic
	Full gene overlap	Coding exon affected, partial	UTR overlap	Intron overlap		
Deletions	654 (631)	1,093 (1,031)	315 (290)	7,319 (6,481)	9,381 (8,433)	12,644 (10,386)
Tandem duplications	2 (2)	7 (6)	9 (5)	197 (62)	215 (75)	286 (76)
Mobile element insertions	—	3 (2)	36 (26)	1,304 (97)	1,348 (112)	4,023 (758)
Novel sequence insertions	—	—	2 (2)	49 (49)	51 (51)	77 (77)
Sum	656 (633)	1,119 (1,040)	351 (309)	8,869 (6,689)	10,995 (8,671)	17,030 (11,280)

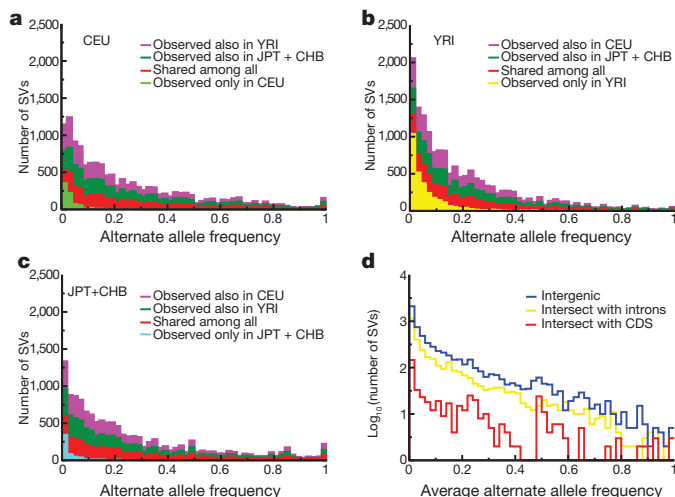
Figures in parentheses indicate numbers of validated SVs per category. We inferred gene overlap with Gencode gene annotation<sup>44</sup>.

Table 9). We related gene disruptions to gene functions, and observed significant enrichments for several functional categories, including cell defence and sensory perception (Supplementary Table 10). High levels of structural variation, including copy number variation, were described previously for both processes<sup>15,22,37</sup>. These SVs might be maintained in the population by selection for the purpose of functional redundancy. Whereas most SVs intersecting with genes were deletions, several validated tandem duplications and MEIs also intersected with coding sequences (Table 2).

### Population genetic properties of deletions

We next sought to generate genotypes for deletions discovered in the 1000GP data, both to facilitate population genetics analyses and to make our SV set amenable to association studies in the form of a reference genotype set. In this regard, the Genome STRiP genotyping method was developed (Handsaker, R. E., Korn, J. M., Nemesh, J. and McCarroll, S. A., unpublished results), a method combining information from RD, RP, SR and haplotype features of population-scale sequence data for genotyping (Methods, Supplementary Text). Using this approach we generated genotypes for 13,826 autosomal deletions in 156 individuals. The genotypes displayed 99.1% concordance with CGH array-based<sup>1</sup> genotypes (available for 1,970 of the deletions), indicating high genotyping accuracy.

Figure 3 presents allele frequency analyses based on these genotypes. As expected, common polymorphisms (minor allele frequency (MAF) > 5%) were typically shared across populations, whereas rare alleles were frequently observed in only one population (Fig. 3a–c). We observed several candidates for monomorphic deletions (that is, genomic segments putatively deleted in all individuals), explainable by rare insertions present in the reference genome or by remaining genotyping inaccuracies (Supplementary Text).



**Figure 3 | Analysis of deletion presence and absence in three populations.** a–c, Deletion allele frequencies and observed sharing of alleles across populations, displayed for deletions discovered in the CEU (a), YRI (b) and JPT + CHB (c) population samples in terms of stacked bars. d, Allele frequency spectra for deletions intersecting with intergenic (blue), intronic (yellow) and protein-coding sequences (red).

Next we assessed the allele frequencies of gene deletions. Similar to a recent array-based study<sup>1</sup>, we observed a depletion of high-frequency alleles among deletions intersecting with protein-coding sequence compared to other deletions ( $P = 2.2 \times 10^{-16}$ ; KS test), consistent with purifying selection keeping most gene deletions at low frequency (Fig. 3d). Nonetheless, several coding sequence deletions were observed with high allele frequency (>80%). Most of these occurred in regions annotated as segmental duplications, consistent with lessened evolutionary constraint in functionally redundant gene categories<sup>22</sup>. Intriguingly, common gene deletions also affected many unique genes with no obvious paralogues. We further analysed the abundance of gene deletions in different populations and observed highly differentiated loci, albeit with no statistically significant relationship between differentiation and particular categories of gene overlap, that is, intronic versus exonic (Supplementary Text).

By comparing deletion genotypes with genotypes of nearby SNPs, we found, consistent with earlier studies<sup>1,13,38</sup>, that deletions in genomic regions accessible to short read sequencing display extensive linkage disequilibrium (LD) with SNPs. Most common deletions (81%) had one or more SNPs with which they are strongly correlated ( $r^2 > 0.8$ ; Supplementary Fig. 7). This indicates that many deletions mapped in our study will be identifiable through tagging SNPs in future studies (Supplementary Text). On the other hand, a fifth of the genotyped deletions were not tagged by HapMap SNPs (a figure similar to the fraction of SNPs that are not tagged by HapMap SNPs<sup>39</sup>), implying that these SVs should be genotyped directly in association studies. Furthermore, the LD properties of complex SVs (for example, multiallelic SVs) have not yet been fully ascertained as methods for genotyping such SVs with similar accuracy are still being developed.

### SV formation mechanism analysis

Nucleotide resolution breakpoint information enables inference of SV formation mechanisms<sup>15,22</sup>. Recent studies broadly distinguished between several germline rearrangement classes, some of which may comprise more than one SV formation mechanism<sup>15,22,40,41</sup>: non-allelic homologous recombination (NAHR), associated with long sequence similarity stretches around the breakpoints; rearrangements in the absence of extended sequence similarity (abbreviated as ‘non-homologous’ or NH), associated with DNA repair by non-homologous end-joining (NHEJ) or with microhomology-mediated break-induced replication (MMBIR); the shrinking or expansion of variable number of tandem repeats (VNTRs), frequently involving simple sequences, by slippage; and MEIs. We distinguished among the classes NAHR, NH, VNTR and MEI by examining the breakpoint junction sequences of SVs that had initially been discovered as deletions or tandem duplications relative to a human reference.

We first compared these SVs to orthologous primate genomic regions to distinguish deletions from insertions/duplications with respect to reconstructed ancestral loci using the BreakSeq classification approach<sup>41</sup>. This analysis showed that of the 11,254 nucleotide-resolution SVs discovered as deletions relative to a human reference, 21% actually represented insertions and 2% represented tandem duplications relative to the putative ancestral genome. Of the remaining SVs, 60% were classified as deletions relative to ancestral sequence, whereas the ancestral state of 17% was undetermined. By comparison,

out of 160 nucleotide-resolution SVs identified as tandem duplications relative to the reference genome, 91.6% were classified as duplications relative to the ancestral genome, whereas the ancestral state of 8.4% remained undetermined (Supplementary Text). Our breakpoint analysis revealed that 70.8% of the deletions and 89.6% of the insertions exhibited breakpoint microhomology/homology ranging from 2–376 bp in size, with distribution modes of 2 bp (attributable to NH) and 15 bp (attributable to MEI), respectively (Fig. 4a, Supplementary Text). As expected<sup>40</sup>, a small portion of the deletions (16.1%) displayed non-template inserted sequences at their breakpoint junctions. By comparison, the tandem duplications showed extensive stretches displaying  $\geq 95\%$  sequence identity at the breakpoints linearly correlating in length with SV size (Fig. 4a). In addition, most tandem duplications displayed 2–17 bp of microhomology at the breakpoint junctions (Supplementary Text).

We subsequently applied BreakSeq<sup>41</sup> to infer formation mechanisms for all SVs classified with regard to ancestral state. Using BreakSeq, we inferred NH as the dominating deletion mechanism, and MEI as the dominating insertion mechanism (Fig. 4b, c and Supplementary Table 11). Furthermore, an abundance of microhomology at tandem duplication breakpoints suggested frequent formation of this SV class by a rearrangement process acting in the absence of homology (NH). When relating SV formation to the variant size spectrum, we observed marked insertion peaks for MEIs at 300 bp, corresponding to *Alu* elements, and at 6 kb, corresponding to the L1 class of long interspersed elements (LINEs) (Fig. 4c). By comparison, NH- and NAHR-based mechanisms occurred across a wide size range, whereas VNTR expansion/shrinkage, consistent with earlier findings<sup>1</sup>, led to relatively small SV sizes (Fig. 4c, d).

Furthermore, when displaying the genomic distribution of SVs (Fig. 5a), we observed a notable clustering of SVs into ‘SV hotspots’. We analysed this clustering in detail by examining the distribution of non-overlapping, adjacent SVs, and observed a marked clustering of

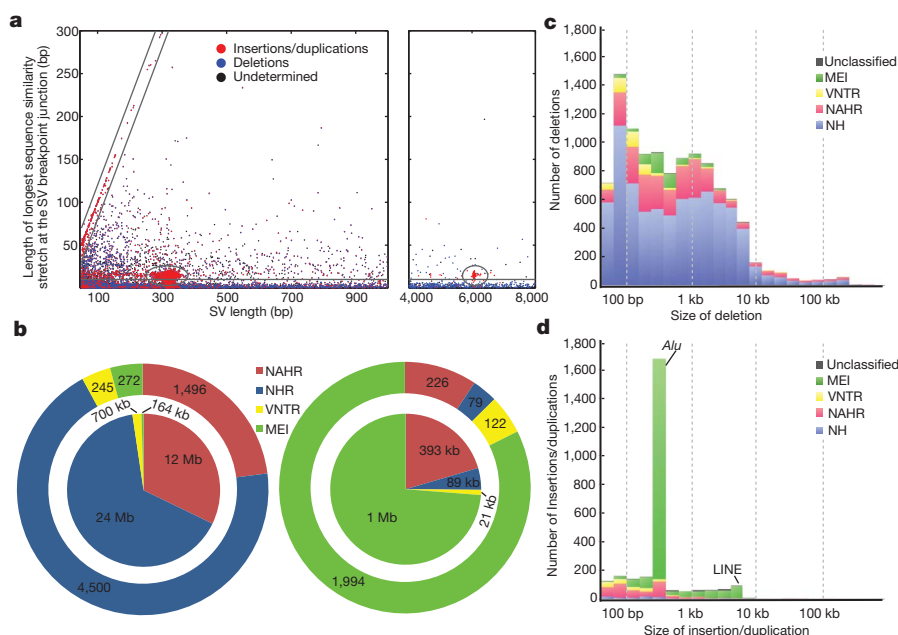
SVs formed by NAHR, VNTR and NH, respectively, a signal extending to hundreds of kilobases (Fig. 5b). The clustering was influenced by an abundance of VNTR near the centromeres<sup>41</sup> and NAHR near the telomeres (Fig. 5a). A significant enrichment of NAHR near recombination hotspots ( $P = 1.3 \times 10^{-15}$ ) and segmental duplications ( $P = 3.1 \times 10^{-17}$ ) further contributed to the clustering (Supplementary Table 13).

To further explore this clustering we devised a segmentation approach for predicting SV hotspots (Methods), which yielded a map of 51 putative SV hotspots (Supplementary Table 14). Most of the hotspots (80%) mainly comprised SVs originating from a single formation mechanism (Fig. 5c). Most of these corresponded to NAHR hotspots, although hotspots dominated by NH and VNTR were also evident. These observations indicate that SV formation is frequently associated with the locus-specific propensity for genomic rearrangement.

## Conclusions and discussion

By generating an SV set of unprecedented size along with breakpoint assemblies and reference genotypes, we demonstrate the suitability of population-scale sequencing for SV analysis. Nucleotide resolution data allow the construction of reference data sets and make SVs readily assessable across different experimental platforms using genotyping approaches. Our fine-scale map enabled us to examine the functional impact of SVs, as exemplified by the set of gene disruption variants we reported, which will be of value for genome and exome sequencing studies.

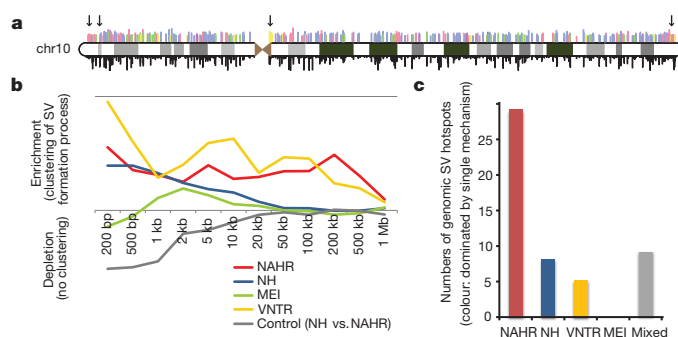
Our map further enabled us to examine size spectra of SV formation mechanisms and led us to identify genomic SV hotspots that are commonly dominated by a single formation mechanism. Recurrent rearrangements, implicated in genomic disorders, are hypothesized to be associated with local genome architecture<sup>42</sup>, for example, with segmental duplications that facilitate NAHR. Also, DNA rearrangement in the absence of homology, that is, MMBIR, has been implicated



**Figure 4 | Contribution of SV formation mechanisms to the SV size spectrum.** **a**, Breakpoint junction homology/microhomology length plotted as a function of SV size for SVs originally identified as deletions compared to a human reference. Dots are coloured according to the SVs' classification as deletions, insertions/duplications, or 'undetermined' relative to inferred ancestral genomic loci. Gray lines mark groups of SVs likely formed by a common formation mechanism. The diagonal highlights tandem duplications (and few reciprocal deletion events), in which the length of the duplicated sequence correlates linearly with the length of the longest breakpoint junction sequence identity stretch. The ellipses indicate MEIs, that is, *Alu* (~300 bp) and

L1 (~6 kb) insertions, associated with target site duplications of up to 28 bp in size at the breakpoints. The horizontal group corresponds mostly to NH-associated deletions with <10 bp microhomology at the breakpoints. The remaining (ungrouped) SVs comprise truncated MEIs, VNTR expansion and shrinkage events, as well as NAHR-associated deletions and duplications. **b**, Relative contributions of SV formation mechanisms in the genome. Numbers of SVs are displayed on the outer pie chart and affected base pairs on the inner. Left panel, SVs classified as deletions relative to ancestral loci. Right panel, SVs classified as insertions/duplications. **c**, Size spectra of deletions classified relative to ancestral loci. **d**, Size spectra of insertions/duplications.





**Figure 5 | Mapping hotspots of SV formation in the genome.** **a**, Distribution of SVs on chromosome 10 ('chr10'). Above the ideogram, coloured bars indicate SV formation mechanisms (same colour scheme as in **(b)** and **(c)**); bar lengths relate to the logarithm of SV size. Below the ideogram, bar lengths are directly proportional to allele frequencies. Arrows indicate an SV hotspot near the centromere underlying mainly VNTR and several hotspots near the telomeres underlying mainly NAHR events. **b**, Enrichment of SVs inferred to be formed by the same formation mechanism for different genomic window sizes. Displayed is an enrichment of nearby, non-overlapping SVs formed by the same mechanism relative to an SV set where mechanism assignments are shuffled randomly. **c**, SV hotspots are mostly dominated by a single formation mechanism. Coloured bars depict numbers of SV hotspots in which at least 50% of the variants were inferred to be formed by a single formation mechanism. The average abundance of NAHR-classified SVs in NAHR hotspots was 70% (compared with 77% for VNTR-hotspots; 69% for NH). The grey bar ('mixed') corresponds to SV hotspots with no single mechanism dominating.

in recurrent SV formation<sup>8,43</sup>. In this regard, we noticed that out of the hotspots we report, six fall into critical regions of known genetic disorders associated with recurrent *de novo* deletions, including Miller-Dieker syndrome and Leri-Weill dyschondrosteosis (Supplementary Table 14). Irrespective of potential disease relevance, or inferred mechanism of formation, our analysis revealed a map of SV hotspots that may constitute local centres of *de novo* SV formation, consistent with the concept that local genome architecture contributes to genomic instability<sup>42</sup>.

Our study focused on characterizing deletions, which are often associated with disease<sup>9</sup>. Facilitated by ancestral analyses of SV loci, we also characterized insertions and tandem duplications, albeit in less detail than deletions. Companion papers with more detailed analyses of MEIs and copy number variation within segmental duplications are published elsewhere (Stewart, C. *et al.*, unpublished results, and ref. 34). Of note, most SV discovery methods depend on mapping reads onto their genomic locus of origin, that is, the 'accessible' fraction of the genome, a fraction lessened in segmental duplications that are of high interest to SV analysis. Nonetheless, owing to the abilities of SV discovery methods in detecting SVs in these regions and in interpreting reads with multiple mapping positions, the 'accessible' fraction of the genome is higher for SVs than for SNPs<sup>16</sup>. In the future, sequencing technologies generating longer DNA reads will increase the accessible genome, and will enable the assessment of SVs embedded in long repeat structures, such as balanced inversions.

Our SV resource will enable the discovery, genotyping and imputation of SVs in larger cohorts. Numerous genomes will be sequenced in the coming months to facilitate disease association studies. Systematic characterization of SVs in these genomes will benefit from the concepts and data sets presented here.

## METHODS SUMMARY

**Samples.** Whole genome sequencing data for 179 unrelated individuals and six individuals from parent-offspring trios were obtained as part of the 1000GP. These data were generated with Illumina/Solexa, Roche/454 and Life Technologies/SOLiD sequencing technology platforms.

**SV discovery and breakpoint assembly.** The SV discovery methods we applied comprised six RP, four RD, three SR, four AS, and two PD based methods. TIGRA (Chen, L. *et al.*, unpublished results) was used for targeted breakpoint assembly.

**Experimental validation.** We validated SV calls by PCR, array CGH and SNP microarrays, targeted assembly, and custom microarray-based sequence capture. PCR was performed in various different laboratories<sup>33</sup>. CGH analysis was performed based on tiling array data provided by the Genome Structural Variation Consortium (ArrayExpress: E-MTAB-40), and SNP array analysis based on data obtained from the International HapMap Consortium (<http://hapmap.ncbi.nlm.nih.gov>).

**Genotyping.** Genome STRiP (Handsaker, R. E., Korn, J. M., Nemesh, J. and McCarroll, S. A., unpublished results) was used for deletion genotyping in low-coverage sequence data. Initial genotype likelihoods were derived with a Bayesian model and imputation into a SNP genotype reference panel from the HapMap<sup>39</sup> (Hapmap3r2) was achieved with Beagle (v3.1; <http://faculty.washington.edu/browning/beagle/beagle.html>).

**SV formation mechanism analysis.** SV breakpoints mapped at nucleotide resolution were analysed with BreakSeq<sup>41</sup> to classify SVs relative to putative ancestral loci and to infer SV formation mechanisms. SV hotspots were mapped with custom Perl and R scripts.

Received 19 August; accepted 26 November 2010.

- Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
- Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
- Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).
- Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
- McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41**, 1223–1227 (2009).
- Craddock, N. *et al.* Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–720 (2010).
- McCarroll, S. A. *et al.* Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nature Genet.* **40**, 1107–1112 (2008).
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nature Rev. Genet.* **10**, 551–564 (2009).
- Stankiewicz, P. & Lupski, J. R. Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* **61**, 437–455 (2010).
- Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
- lafrate, A. J. *et al.* Detection of large-scale variation in the human genome. *Nature Genet.* **36**, 949–951 (2004).
- Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
- McCarroll, S. A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genet.* **40**, 1166–1174 (2008).
- Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nature Genet.* **37**, 727–732 (2005).
- Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
- Alkan, C. *et al.* Personalized copy number and segmental duplication maps using next-generation sequencing. *Nature Genet.* **41**, 1061–1067 (2009).
- Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods* **6**, 677–681 (2009).
- Hormozdiari, F., Alkan, C., Eichler, E. E. & Sahinalp, S. C. Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome Res.* **19**, 1270–1278 (2009).
- Medvedev, P., Stanciu, M. & Brudno, M. Computational methods for discovering structural variation with next-generation sequencing. *Nature Methods* **6**, S13–S20 (2009).
- McKernan, K. J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res.* **19**, 1527–1541 (2009).
- Chiang, D. Y. *et al.* High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nature Methods* **6**, 99–103 (2009).
- Kidd, J. M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
- Lee, S., Cheran, E. & Brudno, M. A robust framework for detecting structural variations in a genome. *Bioinformatics* **24**, i59–i67 (2008).
- Pang, A. W. *et al.* Towards a comprehensive structural variation map of an individual human genome. *Genome Biol.* **11**, R52 (2010).
- Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
- Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
- Yoon, S., Xuan, Z., Makarov, V., Ye, K. & Sebat, J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* **19**, 1586–1592 (2009).
- Mills, R. E. *et al.* An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* **16**, 1182–1190 (2006).

29. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
30. Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome Res.* **19**, 1117–1123 (2009).
31. Hajirasouliha, I. *et al.* Detection and characterization of novel sequence insertions using paired-end next-generation sequencing. *Bioinformatics* **26**, 1277–1283 (2010).
32. Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**, 311–317 (2010).
33. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
34. Sudmant, P. H. *et al.* Diversity of human copy number variation and multicopy genes. *Science* **330**, 641–646 (2010).
35. Willer, C. J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genet.* **41**, 25–34 (2008).
36. Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol.* **5**, e254 (2007).
37. Hasin-Brumshtein, Y., Lancet, D. & Olender, T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet.* **25**, 178–184 (2009).
38. Hinds, D. A., Kloek, A. P., Jen, M., Chen, X. & Frazer, K. A. Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nature Genet.* **38**, 82–85 (2006).
39. Altshuler, D. M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
40. Conrad, D. F. *et al.* Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nature Genet.* **42**, 385–391 (2010).
41. Lam, H. Y. *et al.* Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nature Biotechnol.* **28**, 47–55 (2010).
42. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**, 417–422 (1998).
43. Lee, J. A., Carvalho, C. M. & Lupski, J. R. A. DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007).
44. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (suppl. 1), S4 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to acknowledge C. Hardy, R. Smith, A. De Witte and S. Giles for their assistance with validation. M.A.B.'s group was supported by a grant from the National Institutes of Health (R01 GM59290) and G.T.M.'s group by grants R01 HG004719 and RC2 HG005552, also from the NIH. J.O.K.'s group was supported by an Emmy Noether Fellowship of the German Research Foundation (Deutsche Forschungsgemeinschaft). J.W.'s group was supported by the National Basic Research

Program of China (973 program no. 2011CB809200), the National Natural Science Foundation of China (30725008; 30890032; 30811130531; 30221004), the Chinese 863 program (2006AA022177; 2006AA022334; 2006AA02A302; 2009AA022707), the Shenzhen Municipal Government of China (grants JC200903190767A; JC200903190772A; ZYC200903240076A; CXB200903110066A; ZYC200903240077A; ZYC200903240076A and ZYC200903240080A) and the Ole Rømer grant from the Danish Natural Science Research Council. E.E.E.'s group was supported by grants P01 HG004120 and U01 HG005209 from the National Institutes of Health. C.L.'s group was supported by grants from the National Institutes of Health: P41 HG004221, R01 GM081533 and U01 HG005209 and X.S. was supported by a T32 fellowship award from the NIH. We thank the Genome Structural Variation Consortium (<http://www.sanger.ac.uk/humgen/cnv/42mio/>) and the International HapMap Consortium for making available microarray data. The authors acknowledge the individuals participating in the 1000 Genomes Project by providing samples, including the Yoruba people of Ibadan, Nigeria, the community at Beijing Normal University, the people of Tokyo, Japan, and the people of the Utah CEPH community. Furthermore, we thank R. Durbin and L. Steinmetz for comments on the manuscript.

**Author Contributions** The authors contributed this study at different levels, as described in the following. SV discovery: K.W., C.S., R.E.H., K.C., C.A., A.A., S.C.Y., R.K.C., A.C., Y.F., I.H., F.H., Z.I., D.K., R.Li, Y.L., C.L., R.Lu, X.J.M., H.E.P., L.D., G.T.M., J.S., Ju.W., Ka.Y., Ke.Y., E.E.E., M.B.G., M.E.H., S.A.M. and J.O.K. SV validation: R.E.M., K.W., K.C., A.A., S.C.Y., F.G., M.K.K., J.K., J.N., A.E.U., X.S., A.M.S., J.A.W., Y.Z., Z.D.Z., M.A.B., J.S., M.S., M.E.H., C.L. and J.O.K. SV genotyping: K.W., R.E.H., J.K., J.N., M.E.H. and S.A.M. Data analysis: R.E.M., C.S., C.A., A.A., R.E.H., K.C., S.C.Y., R.K.C., A.C., D.F.C., Y.F., F.H., L.M.I., Z.I., J.M.K., M.K.K., S.K., J.K., E.K., D.K., H.Y.K.L., J.L., R.Li, Y.L., C.L., R.Luo, X.J.M., J.N., H.E.P., T.R., A.S., X.S., M.P.S., J.A.W., Ji.W., Y.Z., Z.D.Z., M.A.B., L.D., G.T.M., G.M., J.S., M.S., Ju.W., Ka.Y., Ke.Y., E.E.E., M.B.G., M.E.H., C.L., S.A.M. and J.O.K. Preparation of manuscript display items: R.E.M., K.W., C.S., C.A., A.A., R.E.H., S.C.Y., L.M.I., S.K., E.K., M.K.K., X.J.M., X.S., J.A.W., M.B.G., S.A.M. and J.O.K. Co-chairs of the Structural Variation Analysis group: E.E.E., M.E.H. and C.L. The following equally contributed to directing the described analyses and participating in the design of the study and should be considered joint senior authors: E.E.E., M.B.G., M.E.H., C.L., S.A.M. and J.O.K. The manuscript was written by the following authors: R.E.M. and J.O.K.

**Author Information** Data sets described here can be obtained from the 1000 Genomes Project website at [www.1000genomes.org](http://www.1000genomes.org) (July 2010 Data Release). Individual SV discovery methods can be obtained from sources mentioned in Supplementary Table 2, or upon request from the authors. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.O.K. ([jan.korbel@embl.de](mailto:jan.korbel@embl.de)).

# An actively accreting massive black hole in the dwarf starburst galaxy Henize 2-10

Amy E. Reines<sup>1</sup>, Gregory R. Sivakoff<sup>1</sup>, Kelsey E. Johnson<sup>1,2</sup> & Crystal L. Brogan<sup>2</sup>

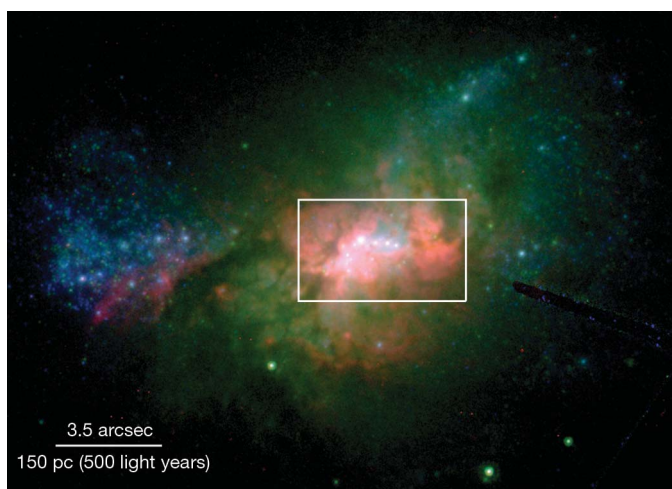
Supermassive black holes are now thought to lie at the heart of every giant galaxy with a spheroidal component, including our own Milky Way<sup>1,2</sup>. The birth and growth of the first ‘seed’ black holes in the earlier Universe, however, is observationally unconstrained<sup>3</sup> and we are only beginning to piece together a scenario for their subsequent evolution<sup>4</sup>. Here we report that the nearby dwarf starburst galaxy Henize 2-10 (refs 5 and 6) contains a compact radio source at the dynamical centre of the galaxy that is spatially coincident with a hard X-ray source. From these observations, we conclude that Henize 2-10 harbours an actively accreting central black hole with a mass of approximately one million solar masses. This nearby dwarf galaxy, simultaneously hosting a massive black hole and an extreme burst of star formation, is analogous in many ways to galaxies in the infant Universe during the early stages of black-hole growth and galaxy mass assembly. Our results confirm that nearby star-forming dwarf galaxies can indeed form massive black holes, and that by implication so can their primordial counterparts. Moreover, the lack of a substantial spheroidal component in Henize 2-10 indicates that supermassive black-hole growth may precede the build-up of galaxy spheroids.

The starburst in Henize 2-10, a relatively nearby (9 megaparsecs,  $\sim 30$  million light years) blue compact dwarf galaxy, has attracted the attention of astronomers for decades<sup>6–10</sup>. Stars are forming in Henize 2-10 at a prodigious rate<sup>8,11,12</sup> that is ten times that of the Large Magellanic Cloud<sup>13</sup> (a satellite galaxy of the Milky Way), despite the fact that both of these dwarf galaxies have similar stellar masses<sup>14–16</sup> and neutral hydrogen gas masses<sup>7,17</sup>. Most of the star formation in Henize 2-10 is concentrated in a large population of very massive and dense ‘super-star clusters’, the youngest having ages of a few million years and masses of one hundred thousand times the mass of the Sun<sup>6</sup>. The main optical body of the galaxy has an extent less than a kiloparsec ( $\sim 3,000$  light-years) in size and has a compact irregular morphology typical of blue compact dwarfs (Fig. 1).

We observed Henize 2-10 at centimetre radio wavelengths with the Very Large Array (VLA) and in the near-infrared with the Hubble Space Telescope (HST) as part of a large-scale panchromatic study of nearby dwarf starburst galaxies harbouring infant super-star clusters<sup>18–20</sup>. A comparison between the VLA and HST observations drew our attention to a compact ( $< 24$  pc  $\times$  9 pc) central radio source located between two bright regions of ionized gas (Fig. 2). These data exclude any association of this central radio source with a visible stellar cluster (Fig. 3; see Supplementary Information for a discussion of the astrometry). Furthermore, the radio emission from this source has a significant non-thermal component ( $\alpha \approx -0.4$ ,  $S_\nu \propto \nu^\alpha$  where  $S_\nu$  is the flux density at frequency  $\nu$ ) between 4.9 GHz and 8.5 GHz, as noted in previous studies of the galaxy<sup>9</sup>. An archival observation of Henize 2-10 taken with the Chandra X-ray Observatory reveals that a point source with hard X-ray emission is also coincident (to within the position uncertainty) with the central non-thermal radio source<sup>10</sup> (see Supplementary Information). Typically, even powerful non-nuclear radio and X-ray sources (for example, supernova remnants and active X-ray binaries)

are at least an order of magnitude less luminous than the central source in Henize 2-10 (see Supplementary Information). In contrast, the radio and hard X-ray luminosities of the central source in Henize 2-10, as well as their ratio, are similar to known low-luminosity active galactic nuclei powered by accretion onto a massive black hole<sup>21</sup>.

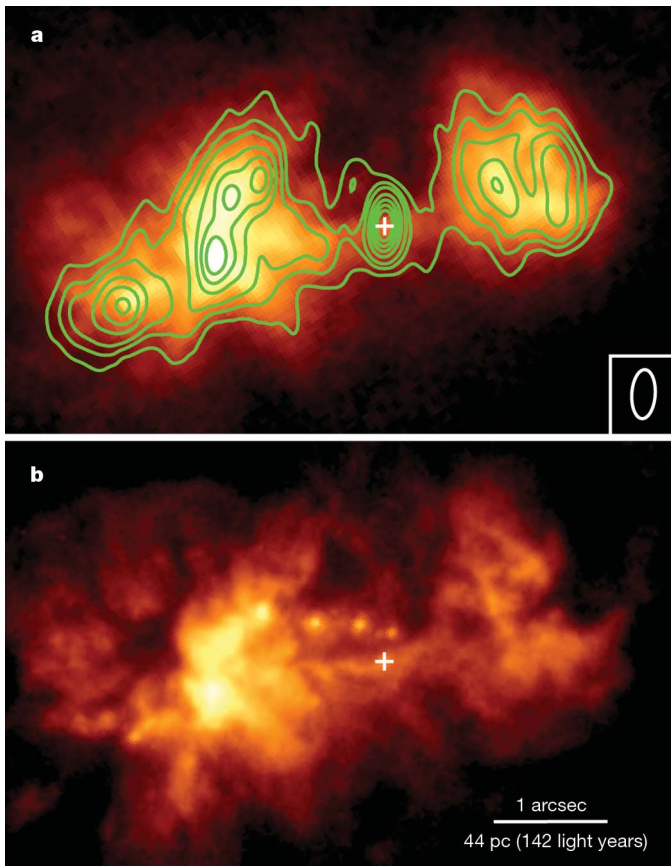
The central, compact, non-thermal radio source in Henize 2-10 is also coincident with a local peak in Pa $\alpha$  and H $\alpha$  emission and appears to be connected to a thin quasi-linear ionized structure between two bright and extended regions of ionized gas. This morphology is tantalizingly suggestive of outflow (Fig. 2). Although we cannot conclusively determine whether or not this linear structure is physically connected to the brightest emitting regions with the data in hand, ground-based spectroscopic observations<sup>22</sup> confirm a coherent velocity gradient along the entire ionized gas structure seen in Fig. 2, consistent with outflow or rotation. Moreover, a comparison between the central velocity of this ionized gas structure and the systemic velocity of the galaxy—derived



**Figure 1 | Henize 2-10.** Henize 2-10 is a blue compact dwarf galaxy hosting a concentrated region of extreme star formation. Using H $\alpha$  (ref. 8) and 24  $\mu$ m (ref. 11) fluxes from the literature, we estimate a star-formation rate<sup>12</sup> of  $1.9 M_\odot \text{ yr}^{-1}$ , assuming that all of the emission is from the starburst and that the contribution from the active nucleus is negligible. We estimate that Henize 2-10 has a stellar mass of  $3.7 \times 10^9 M_\odot$  from the integrated 2MASS K $_s$ -band flux<sup>14,15</sup>. Neutral hydrogen observations of Henize 2-10 indicate a solid-body rotation curve typical of dwarf galaxies with a maximum projected rotational velocity of  $39 \text{ km s}^{-1}$  relative to the systemic velocity of the galaxy<sup>7</sup>. These observations also indicate a dynamical mass of about  $10^{10} M_\odot$  within 2.1 kiloparsecs (ref. 7). The main optical body of the galaxy, shown here, is less than one kiloparsec across. Henize 2-10 shows signs of having undergone an interaction, including tidal-tail-like features in both its gaseous<sup>7</sup> and stellar distributions (seen here). In this three-colour HST image of the galaxy, we show ionized gas emission in red (H $\alpha$ ) and the stellar continuum in green ( $\sim I$ -band,  $0.8 \mu\text{m}$ ) and blue ( $\sim U$ -band,  $0.3 \mu\text{m}$ ). These archival data were taken with Wide Field and Planetary Camera 2 (H $\alpha$ ) and the Advanced Camera for Surveys ( $U$ - and  $I$ -band). The white box indicates the region shown in Figs 2 and 3.

<sup>1</sup>Department of Astronomy, University of Virginia, 530 McCormick Road, Charlottesville, Virginia 22904, USA. <sup>2</sup>National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22904, USA.

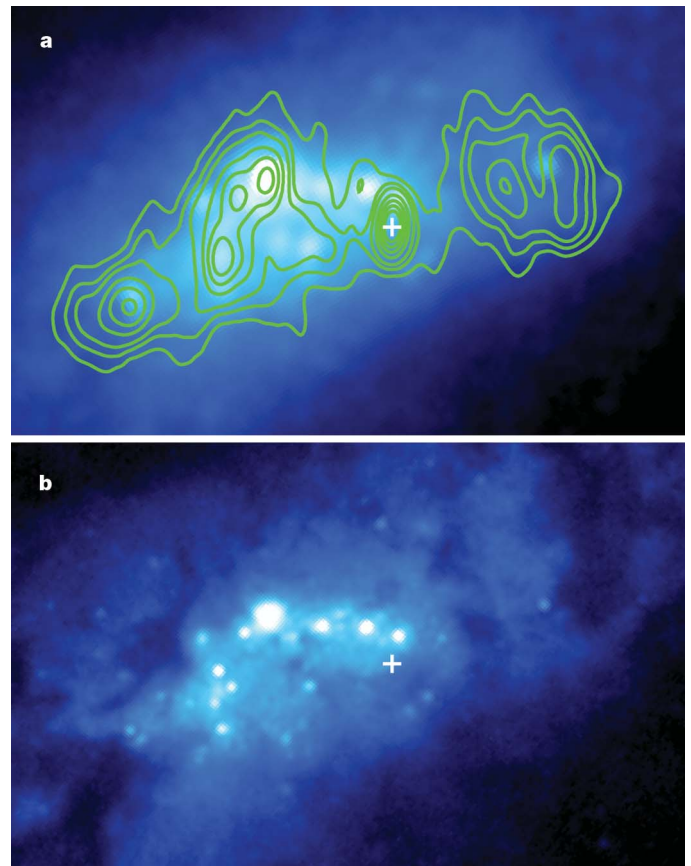




**Figure 2 | The active nucleus and ionized gas in Henize 2-10.** The overall morphology of the radio emission (green contours) in the central region of Henize 2-10 matches that of the ionized gas detected with HST (colour images), suggesting a shared origin. The active nucleus (plus symbol) is detected as a non-thermal nuclear VLA radio source coincident with a Chandra point source with hard X-ray emission. The nuclear source is also coincident with a local peak of ionized gas emission and appears to be connected to the thin quasi-linear structure between the two bright and extended regions of ionized gas: this is suggestive of, although not proof of, outflow. The central source has 4.9 GHz and 8.5 GHz radio luminosities of  $7.4 \times 10^{35} \text{ erg s}^{-1}$  and  $1.0 \times 10^{36} \text{ erg s}^{-1}$ , respectively. The X-ray luminosity of the central source in the 2–10 keV band is  $\sim 2.7 \times 10^{39} \text{ erg s}^{-1}$ . The accretion rate of the  $2 \times 10^6 M_{\odot}$  black hole is  $\sim 5 \times 10^{-6} M_{\odot} \text{ yr}^{-1}$  assuming an X-ray bolometric correction of 10 and an accretion efficiency of 0.1. **a**, Narrowband imaging with the Near Infrared Camera and Multi-Object Spectrometer (NICMOS) on the Hubble Space Telescope was used to trace the ionized gas in Henize 2-10 using the Pa $\alpha$  hydrogen recombination line at  $1.87 \mu\text{m}$ . Continuum emission was removed using a neighbouring off-line narrowband filter. VLA 8.5 GHz (3.5 cm) radio contours are over-plotted in green and the active galactic nucleus is marked with a plus symbol. Contour levels are 9, 13, 17, 25, 33, 41 and 49 times the root-mean-square noise ( $12 \mu\text{Jy}$  per beam). The beam is shown in the lower right corner. **b**, Optical narrowband imaging of the H $\alpha$  hydrogen recombination line at  $0.66 \mu\text{m}$  yields a higher-resolution view of the ionized gas in Henize 2-10. The continuum has not been subtracted in this archival image, leaving young star clusters also visible.

from observations of neutral hydrogen gas rotating as a solid body<sup>7</sup>—indicates that the position of the non-thermal radio source is consistent with the dynamical centre of the galaxy.

Compact radio and hard X-ray emission at the centre of a galaxy are generally good indicators of accretion onto a massive black hole<sup>21</sup>, but we have also considered alternative explanations for the data. As discussed at length in the Supplementary Information, it is extremely unlikely that the nuclear source in Henize 2-10 is one or more supernova remnants, more recently created supernovae, stellar-mass black-hole X-ray binaries, or some combination of these phenomena. Briefly, X-ray binaries are too weak in the radio, supernova remnants are too weak in hard X-rays, and young compact radio supernovae are ruled out



**Figure 3 | Young super-star clusters in Henize 2-10.** The overall structure of the radio emission (green contours) differs markedly from the distribution of star clusters in the centre of the galaxy (colour images). In particular, the non-thermal nuclear radio source does not have a detectable counterpart in these broadband continuum images (plus symbol), excluding any association with a visible star cluster. **a**, A near-infrared ( $\sim K$ -band) image of the central region of Henize 2-10 overplotted with the same radio contours as in Fig. 2. HST/NICMOS was used to observe the galaxy through a broadband filter centred at  $2.1 \mu\text{m}$ , which primarily traces the distribution of stellar light. **b**, A higher-resolution view of the star clusters is shown in this archival  $0.8 \mu\text{m}$  ( $\sim I$ -band) broadband image. The field of view is the same as in Fig. 2.

by observations using Very Long Baseline Interferometry<sup>23</sup>. Although it may be possible to account for the radio and X-ray luminosities of the nuclear source with just the right combination of the abovementioned phenomena, the probability of such a coincidence is exceedingly low (see Supplementary Information). On the contrary, the radio and hard X-ray luminosities of the central source in Henize 2-10 are well within the range of known low-luminosity active galactic nuclei<sup>21</sup>.

In addition to ruling out young compact supernovae, the non-detection of the nuclear radio source at higher resolution ( $\sim 0.5 \text{ pc} \times 0.1 \text{ pc}$ ) using Very Long Baseline Interferometry<sup>23</sup> may also seemingly rule out the presence of an actively accreting massive black hole. However, Seyfert nuclei with steep radio spectra ( $\alpha \lesssim -0.5$ ,  $S_{\nu} \propto \nu^{\alpha}$ ) often exhibit this ‘missing flux’ phenomenon when observed at increasingly higher spatial resolution<sup>24</sup>. In these active galactic nuclei, as much as  $\sim 90\%$  of the radio emission is absent on parsec scales and is expected to be dominated by extended low-surface-brightness features on larger scales, such as jets. This is in contrast to Seyferts with flat or positive radio spectra ( $\alpha \gtrsim 0$ ) and elliptical radio galaxies in which the radio emission is concentrated in a compact core. The nuclear radio source in Henize 2-10 has a spectral index ( $\alpha \approx -0.4$ ) similar to Seyfert nuclei that are known to have reduced flux densities on parsec scales. Therefore, we do not consider the non-detection of the nuclear radio source at very high resolution to be incompatible with the presence of an active galactic nucleus in Henize 2-10.

We conclude that an actively accreting massive black hole is the most feasible explanation for the nuclear source in Henize 2-10. The compact radio and hard X-ray luminosities are consistent with the observed correlation for active galactic nuclei and we therefore estimate the mass of the black hole in Henize 2-10 using the so-called “fundamental plane of black hole activity”<sup>23</sup>. This empirical relationship relating black-hole mass to the emitted compact radio and hard X-ray luminosities, spanning nine orders of magnitude in black-hole mass, is given by the equation  $\log L_R = 0.60 \log L_X + 0.78 \log M + 7.33$ , where  $L_R$  is the radio luminosity at 5 GHz in  $\text{erg s}^{-1}$ ,  $L_X$  is the 2–10 keV X-ray luminosity in  $\text{erg s}^{-1}$ , and  $M$  is the mass of the black hole in solar masses,  $M_\odot$ . Using the observed radio luminosity of  $7.4 \times 10^{35} \text{ erg s}^{-1}$  at 4.9 GHz and the X-ray luminosity of  $2.7 \times 10^{39} \text{ erg s}^{-1}$  in the 2–10 keV band, we calculate  $\log(M/M_\odot) = 6.3 \pm 1.1$  for the black hole in Henize 2-10. The region in which the gravitational influence of a one-million-solar-mass black hole dominates that of the host galaxy subtends a very small angle on the sky at the distance of Henize 2-10 ( $< 1$  arcsecond for velocity dispersions  $> 10 \text{ km s}^{-1}$ ). Thus, it is not surprising that kinematic studies of Henize 2-10 have not previously revealed the presence of the black hole at its centre.

Few dwarf galaxies are currently known to host massive black holes<sup>26,27</sup>; however, the discovery of an active nucleus in Henize 2-10 opens up a new realm in which to search for local analogues of primordial black-hole growth (that is, dwarf starburst galaxies). While recent searches<sup>28,29</sup> have revealed growing numbers of nuclear black holes with masses similar to that in Henize 2-10, the host galaxies of these objects have very different properties from that of Henize 2-10. Most notably, they are not actively forming stars and have regular morphologies of disks and spheroids with well-defined optical nuclei<sup>29,30</sup>. Moreover, the majority of the black holes detected in these systems are radiating at high fractions of their Eddington limits<sup>27–29</sup>, suggesting that the black holes are currently undergoing rapid growth. In contrast, the low-luminosity active galactic nucleus in Henize 2-10 is currently radiating significantly below its Eddington limit ( $\sim 10^{-4}$  assuming an X-ray bolometric correction of ten; see Supplementary Information), suggesting a different evolutionary state.

The results presented here have broad implications for our understanding of the evolution of galaxies and their central black holes. The concurrent black-hole growth and extreme starburst in Henize 2-10 probably resembles the conditions in low-mass, high-redshift galaxies during the early phases of galaxy assembly when interactions and mergers were common. Indeed, Henize 2-10 shows signs of having undergone an interaction, including tidal-tail-like features in both its gaseous<sup>7</sup> and stellar distributions (Fig. 1). Additionally, it is intriguing that the massive black hole in Henize 2-10 does not appear to be associated with a bulge, a nuclear star cluster or any other well-defined nucleus. This unusual property may reflect an early phase of black-hole growth and galaxy evolution that has not been previously observed. If so, this implies that primordial seed black holes could have pre-dated their eventual dwellings, thereby constraining theories for the formation mechanisms of massive black holes and galaxies.

Received 1 September; accepted 30 November 2010.

Published online 9 January 2011.

- Magorrian, J. *et al.* The demography of massive dark objects in galaxy centers. *Astron. J.* **115**, 2285–2305 (1998).
- Ghez, A. M. *et al.* Stellar orbits around the galactic center black hole. *Astrophys. J.* **620**, 744–757 (2005).
- Volonteri, M. Formation of supermassive black holes. *Astron. Astrophys. Rev.* **18**, 279–315 (2010).
- Heckman, T. The co-evolution of galaxies and black holes: current status and future prospects. *IAU Symp.* **267**, 3–14 (2010).
- Allen, D. A., Wright, A. E. & Goss, W. M. The dwarf emission galaxy He2–10. *Mon. Not. R. Astron. Soc.* **177**, 91–97 (1976).
- Johnson, K. E., Leitherer, C., Vacca, W. D. & Conti, P. S. Hubble Space Telescope observations of HE 2–10: outflows and young super-star clusters. *Astron. J.* **120**, 1273–1288 (2000).

- Kobulnicky, H. A., Dickey, J. M., Sargent, A. I., Hogg, D. E. & Conti, P. S. Aperture synthesis observations of molecular and atomic gas in the Wolf-Rayet starburst galaxy Henize 2–10. *Astron. J.* **110**, 116–130 (1995).
- Méndez, D. I. *et al.* A complex bipolar outflow in the Wolf-Rayet BCDG He 2–10. *Astron. Astrophys.* **349**, 801–811 (1999).
- Johnson, K. E. & Kobulnicky, H. A. The spectral energy distributions of infant super-star clusters in Henize 2–10 from 7 millimeters to 6 centimeters. *Astrophys. J.* **597**, 923–928 (2003).
- Kobulnicky, H. A. & Martin, C. L. The diffuse and compact X-ray components of the starburst galaxy Henize 2–10. *Astrophys. J.* **718**, 724–738 (2010).
- Engelbracht, C. W. *et al.* Metallicity effects on mid-infrared colors and the 8  $\mu\text{m}$  PAH emission in galaxies. *Astrophys. J.* **628**, L29–L32 (2005).
- Calzetti, D. *et al.* The calibration of mid-infrared star formation rate indicators. *Astrophys. J.* **666**, 870–895 (2007).
- Whitney, B. A. *et al.* Spitzer Sage Survey of the Large Magellanic Cloud. III. Star formation and  $\sim 1000$  new candidate young stellar objects. *Astron. J.* **136**, 18–43 (2008).
- Skrutskie, M. F. *et al.* The Two Micron All Sky Survey (2MASS). *Astron. J.* **131**, 1163–1183 (2006).
- Bell, E. F., McIntosh, D. H., Katz, N. & Weinberg, M. D. The optical and near-infrared properties of galaxies. I. Luminosity and stellar mass functions. *Astrophys. J.* **149** (Supp.), 289–312 (2003).
- van der Marel, R. P., Roeland, P., Alves, D. R., Hardy, E. & Suntzeff, N. B. New understanding of Large Magellanic Cloud structure, dynamics, and orbit from carbon star kinematics. *Astron. J.* **124**, 2639–2663 (2002).
- Kim, S. *et al.* Aperture synthesis mosaic of the Large Magellanic Cloud. *Astrophys. J.* **503**, 674–688 (1998).
- Reines, A. E., Johnson, K. E. & Goss, W. M. Emerging massive star clusters revealed: high-resolution imaging of NGC 4449 from the radio to the ultraviolet. *Astron. J.* **135**, 2222–2239 (2008).
- Reines, A. E., Johnson, K. E. & Hunt, L. K. A new view of the super star clusters in the low-metallicity galaxy SBS 0335–052. *Astron. J.* **136**, 1415–1426 (2008).
- Reines, A. E., Nidever, D. L., Whelan, D. G. & Johnson, K. E. The importance of nebular continuum and line emission in observations of young massive star clusters. *Astrophys. J.* **708**, 26–37 (2010).
- Ho, L. C. Nuclear activity in nearby galaxies. *Annu. Rev. Astron. Astrophys.* **46**, 475–539 (2008).
- Henry, A. L., Turner, J. L., Beck, S. C., Crosthwaite, L. P. & Meier, D. S. Brackett lines from the super star cluster nebulae in HE 2–10. *Astron. J.* **133**, 757–767 (2007).
- Ulvestad, J. S., Johnson, K. E. & Neff, S. G. A VLBI search for radio supernovae in Wolf-Rayet galaxies. *Astron. J.* **133**, 1868–1873 (2007).
- Orienti, M. & Prieto, M. A. Radio structures of the nuclei of nearby Seyfert galaxies and the nature of the missing diffuse emission. *Mon. Not. R. Astron. Soc.* **401**, 2599–2610 (2010).
- Merloni, A., Heinz, S. & di Matteo, T. A fundamental plane of black hole activity. *Mon. Not. R. Astron. Soc.* **345**, 1057–1076 (2003).
- Filippenko, A. V. & Sargent, W. L. W. Discovery of an extremely low luminosity Seyfert nucleus in the dwarf galaxy NGC 4395. *Astrophys. J.* **342**, L11–L14 (1989).
- Barth, A. J., Ho, L. C., Rutledge, R. E. & Sargent, W. L. W. POX 52: a dwarf Seyfert 1 galaxy with an intermediate-mass black hole. *Astrophys. J.* **607**, 90–102 (2004).
- Greene, J. E. & Ho, L. C. A new sample of low-mass black holes in active galaxies. *Astrophys. J.* **670**, 92–104 (2007).
- Barth, A. J., Greene, J. E. & Ho, L. C. Low-mass Seyfert 2 galaxies in the Sloan Digital Sky Survey. *Astron. J.* **136**, 1179–1200 (2008).
- Greene, J. E., Ho, L. C. & Barth, A. J. Black holes in pseudobulges and spheroidals: a change in the black hole-bulge scaling relations at low mass. *Astrophys. J.* **688**, 159–179 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** A.E.R. is grateful for many discussions on this work, in particular with M. Whittle, J. Ulvestad, M. Goss, S. Kannappan, J. Greene, R. Hickox, A. Evens, R. O’Connell, R. Chevalier, A. Seth, E. Gallo, S. Ransom, L. Hunt and J. Simon. A.E.R. acknowledges support from a NASA Earth and Space Science Fellowship, and the University of Virginia through a Governor’s Fellowship and a Dissertation Acceleration Fellowship. G.R.S. acknowledges support for this work by NASA through the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA. K.E.J. acknowledges support from the NSF through a CAREER award and the David and Lucile Packard Foundation through a Packard Fellowship. Support was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This research has made use of data obtained from the Hubble Space Telescope and Chandra X-ray Observatory Data Archives.

**Author Contributions** A.E.R. reduced the HST/NICMOS images, synthesized the multi-wavelength data, and led the analysis, interpretation, and writing of the paper. G.R.S. analysed the Chandra data, and helped with the interpretation and writing of the paper. K.E.J. led the HST/NICMOS and VLA proposals. K.E.J. and C.L.B. reduced and analysed the VLA data. All authors discussed the results and presentation of the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.E.R. ([areines@virginia.edu](mailto:areines@virginia.edu)).



# Entanglement in a solid-state spin ensemble

Stephanie Simmons<sup>1</sup>, Richard M. Brown<sup>1</sup>, Helge Riemann<sup>2</sup>, Nikolai V. Abrosimov<sup>2</sup>, Peter Becker<sup>3</sup>, Hans-Joachim Pohl<sup>4</sup>, Mike L. W. Thewalt<sup>5</sup>, Kohei M. Itoh<sup>6</sup> & John J. L. Morton<sup>1,7</sup>

Entanglement is the quintessential quantum phenomenon. It is a necessary ingredient in most emerging quantum technologies, including quantum repeaters<sup>1</sup>, quantum information processing<sup>2</sup> and the strongest forms of quantum cryptography<sup>3</sup>. Spin ensembles, such as those used in liquid-state nuclear magnetic resonance<sup>4,5</sup>, have been important for the development of quantum control methods. However, these demonstrations contain no entanglement and ultimately constitute classical simulations of quantum algorithms. Here we report the on-demand generation of entanglement between an ensemble of electron and nuclear spins in isotopically engineered, phosphorus-doped silicon. We combined high-field (3.4 T), low-temperature (2.9 K) electron spin resonance with hyperpolarization of the <sup>31</sup>P nuclear spin to obtain an initial state of sufficient purity to create a non-classical, inseparable state. The state was verified using density matrix tomography based on geometric phase gates, and had a fidelity of 98% relative to the ideal state at this field and temperature. The entanglement operation was performed simultaneously, with high fidelity, on 10<sup>10</sup> spin pairs; this fulfils one of the essential requirements for a silicon-based quantum information processor.

Most quantum information processing algorithms applied to spin ensembles have been implemented in a regime of weak spin polarization. However, owing to the very low purity of the states used, any exponential enhancement offered by quantum mechanics disappears when the scaling of total resources is considered. Highly mixed, or weakly initialized, ensembles are often interpreted as the sum of a perfectly mixed component (given by a normalized identity matrix in the density matrix representation) and a small amount,  $\varepsilon$ , of a pure component,  $\rho_0$ ; thus,  $\rho_{\text{true}} = (1 - \varepsilon)\hat{I}/d + \varepsilon\rho_0$ , where  $d$  is the dimensionality of the state. The  $\hat{I}$  component is invariant under unitary operations and is not directly observable by magnetic resonance, which produces measurements of the population differences across allowed electron and nuclear spin transitions. It is therefore straightforward to ignore the maximally mixed component: this approach is called the 'pseudo-pure approximation'<sup>6</sup>.

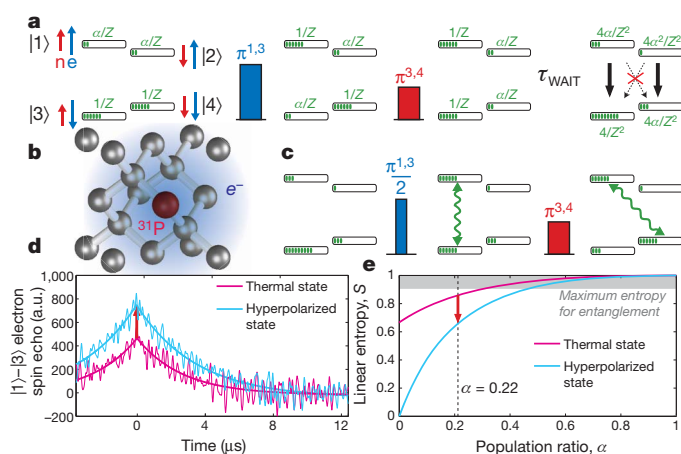
There are a number of entanglement witnesses or monotones that can distinguish entangled states from (classical) separable ones. A widely used test is the positive partial transpose (PPT) criterion, which is both a necessary and a sufficient test of entanglement for two coupled, spin-1/2 particles<sup>7,8</sup>. Applying this test to the mixed state above,  $\rho_{\text{true}}$ , it can be shown<sup>8</sup> that the minimum value of  $\varepsilon$  which permits the overall state to be entangled is 1/3.

Typical values for  $\varepsilon$  in liquid-state nuclear magnetic resonance and electron spin resonance (using 10-GHz excitation at a temperature of 5 K) are  $\sim 10^{-5}$  and  $\sim 10^{-2}$ , respectively. These values are well below the required threshold for the PPT test. Thus, although experiments performed in this regime provide a valuable test bed for techniques in entanglement generation and detection<sup>9</sup>, the states created are only pseudo-entangled, and are fully separable. (A notable exception was the use of chemical methods to generate highly polarized hydrogen spin pairs<sup>10</sup>, though that is a single-shot experiment with limited scalability.)

To overcome this limit, we require states of higher initial purity and a method to measure the  $\hat{I}$  component of the density matrix.

We follow a hybrid approach, using both the electron spin and the nuclear spin associated with a phosphorus donor in silicon. Isolated donors in isotopically engineered semiconductors are of particular interest as they possess excellent decoherence characteristics (both the electron and the nuclear coherence times,  $T_2$ , exceed seconds<sup>11,12</sup>), can be controlled with high fidelity using microwave and radio-frequency pulses<sup>13,14</sup>, and are promising for integrating quantum technologies into conventional semiconductor devices<sup>15</sup>.

Neglecting the weak polarization of the nuclear spin, the initial state populations are determined by the electron spin Zeeman energy, as shown in Fig. 1a, where  $\alpha = \exp(-g\mu_B B/k_B T)$ ,  $g$  is the electron  $g$ -factor,  $\mu_B$  is the Bohr magneton,  $k_B$  is Boltzmann's constant, and  $B$  and  $T$  are the experimental magnetic field and temperature, respectively. At a high magnetic field (3.4 T) and low temperature (2.9 K), the donor electron spin is thermally polarized to  $\sim 66\%$ ; however, the <sup>31</sup>P nuclear spin, with a much weaker magnetic moment, has only  $\sim 0.04\%$  polarization. Various methods, collectively known as dynamic nuclear



**Figure 1 | Sequences for nuclear spin hyperpolarization and entanglement generation for this coupled  $S = 1/2$ ,  $I = 1/2$  spin system.** **a**, The initial state is at thermal equilibrium, where populations (green) are distributed according to the electron spin ( $e$ ) polarization at this magnetic field and temperature (see text). A pair of applied microwave and radio-frequency  $\pi$  pulses move spin populations to favour the  $|\uparrow\rangle$  nuclear spin ( $n$ ) state. After some time,  $\tau_{\text{WAIT}} \gg T_{1e}$ , there is a significant majority population in state  $|3\rangle$ , or  $|\uparrow\downarrow\rangle$  (where the first and second arrows indicate the nuclear and electron spins, respectively). Nuclear spin and cross-relaxation processes occur on timescales much longer than  $T_{1e}$ . **b**, Illustration of the <sup>28</sup>Si:P coupled spin system. **c**, Starting from the hyperpolarized state in **a**, an electron spin coherence is generated and transformed into the final entangled state, containing a superposition of  $|\uparrow\uparrow\rangle$  and  $|\downarrow\downarrow\rangle$ . **d**, The growth in the electron spin echo intensity measured on the  $|1\rangle$ – $|3\rangle$  transition provides a measure of the population ratio,  $\alpha$ . a.u., arbitrary units. **e**, This hyperpolarization sequence minimizes the linear entropy of the two-spin state for a given value of  $\alpha$ .

<sup>1</sup>Department of Materials, Oxford University, Oxford OX1 3PH, UK. <sup>2</sup>Leibniz-Institut für Kristallzüchtung, 12489 Berlin, Germany. <sup>3</sup>PTB Braunschweig, 38116 Braunschweig, Germany. <sup>4</sup>VITCON Projectconsult GmbH, 07743 Jena, Germany. <sup>5</sup>Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. <sup>6</sup>School of Fundamental Science and Technology, Keio University, Yokohama, 3-14-1 Hiyoshi, 223-8522, Japan. <sup>7</sup>CAESR, Clarendon Laboratory, Oxford University, Oxford OX1 3PU, UK.



polarization<sup>16,17</sup>, exist for indirectly transferring electron spin polarization to the nuclear spin and often exploit cross-relaxation processes involving simultaneous electron and nuclear spin flips. Here we exploit the relative absence of cross-relaxation leading to a substantial difference in the relaxation times of the electron and nuclear spins<sup>13</sup>, to hyperpolarize the nuclear spin rapidly and with high efficiency. This hyperpolarization process is similar to ‘algorithmic cooling’ methods, whereby a particular quantum bit (qubit) relaxes quickly owing to coupling to a heat bath<sup>18</sup>.

Figure 1 illustrates our method for tackling the twin challenges of measuring and minimizing the  $\hat{I}$  component in the density matrix of the coupled electron–nuclear spin system. The hyperpolarization of the nuclear spin can be understood as a SWAP operation (which interchanges the states of two qubits) with the (thermally polarized) electron spin, using a combination of resonant microwave and radio-frequency  $\pi$  pulses. This is followed by a delay  $\tau_{\text{WAIT}}$ , which is substantially longer than the electron spin relaxation time,  $T_{1e}$  (specifically,  $\tau_{\text{WAIT}} \approx 8T_{1e}$ ), during which the electron spin relaxes back to thermal equilibrium. On this timescale, other relaxation processes (such as pure nuclear spin flips or electron–nuclear spin flip-flops) are orders of magnitude slower and can be neglected. The resulting hyperpolarized state is

$$\rho = \frac{4}{Z} (\alpha|1\rangle\langle 1| + \alpha^2|2\rangle\langle 2| + |3\rangle\langle 3| + \alpha|4\rangle\langle 4|)$$

where  $Z = 2(1 + \alpha)$  is a normalizing constant.

Although spin echo sequences can only be used to probe the population differences across energy levels, we can obtain a direct measure of the population ratio,  $\alpha$ , by measuring the electron spin echo amplitude between levels  $|1\rangle$  and  $|3\rangle$  before and after the hyperpolarization sequence, as shown in Fig. 1d. Owing to the enhanced polarization of the nuclear spin, a spin echo measured on this transition increases by a factor of  $2/(\alpha + 1)$  in comparison with the measurement from a fully relaxed thermal state. This measure is strictly conservative: it places a lower bound on the true polarization of the electron, as imperfections such as pulse errors or residual relaxation processes only lead to a lower apparent state purity. Using this measure, we observe an enhancement of the echo intensity by a factor of 1.643(2), corresponding to an upper bound of  $\alpha \leq 0.217(2)$ .

Linear spin entropy (defined as  $\mathcal{N}[1 - \text{Tr}(\rho^2)]/(\mathcal{N} - 1)$  for an  $\mathcal{N}$ -dimensional Hilbert space) is a useful characterization of a state’s purity, and ranges from one, for maximally mixed states, to zero, for pure states. Our hyperpolarization sequence corresponds to a decrease in linear spin entropy, made possible by the open quantum system’s contact with the lattice heat bath (Fig. 1e). Importantly, this approach leads to the minimum possible linear entropy given the electron spin polarization resource and type of relaxation present<sup>18</sup>. Entanglement is maximized in a mixed, two-qubit density matrix by first minimizing the linear entropy and then generating an entangled coherence across the levels with the largest and second-smallest populations<sup>19,20</sup>. Following this strategy, we create an entangled state using a coherence-generating microwave  $\pi^{1,3}/2$  pulse (where the superscript denotes the pair of levels addressed by the pulse) followed by a radio-frequency  $\pi^{3,4}$  pulse (Fig. 1c), yielding the target state:

$$\rho = \frac{1}{2Z^2} \begin{pmatrix} 1+\alpha & 0 & 0 & 1-\alpha \\ 0 & 2\alpha^2 & 0 & 0 \\ 0 & 0 & 2\alpha & 0 \\ 1-\alpha & 0 & 0 & 1+\alpha \end{pmatrix}$$

This density matrix is entangled according to the PPT criterion when  $\alpha \leq 0.432$ ; other preparation methods (such as pseudo-pure state preparation) require substantially higher polarization (Supplementary Information).

Having prepared the initial state and performed an entangling operation, we now use density matrix tomography to extract the final two-spin state. Owing to the weak magnetic moment of nuclear spins

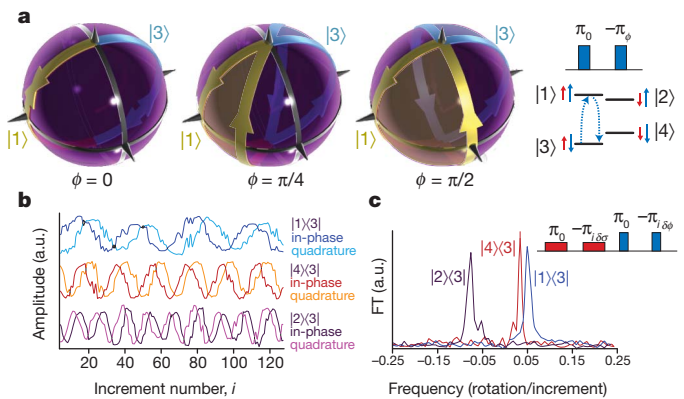
and necessarily low donor concentration in our sample, we are restricted to non-projective measurements of the electron spin ensemble along the  $\sigma_x$  and  $\sigma_y$  Pauli bases, which can be performed selectively on the  $m_I$  state of the nuclear spin (in product operator formalism, these bases can be written as  $S_{x,y}I^z, \beta$ ).

Diagonal elements of the density matrix (corresponding to state populations) are obtained by mapping pairs of population differences into an electron spin echo on the  $|1\rangle$ – $|3\rangle$  transition ( $S_{x,y}I^z$ ). The accurate detection of off-diagonal elements (coherences) is a more elaborate process, made by selectively labelling the coherence between each pair of eigenstates with a distinguishable, time-varying phase<sup>9</sup>. By this process, a particular phase accumulation rate provides the signature of a particular coherence, allowing the off-diagonal elements to be reconstructed from the amplitudes in the Fourier transform of a measured signal.

Here we follow an approach inspired by the Aharonov–Anandan geometric phase gate<sup>21,22</sup> to apply arbitrary phases in a fixed time to the four different eigenstates, and thus separately label each of the possible coherences. We apply two  $\pi$  pulses, along different axes, across a transition between a pair of eigenstates. The phase acquired by each eigenstate is opposite and equal to half the solid angle of its trajectory on the Bloch sphere (Fig. 2a). Thus, applying  $\pi_0^{1,3}$  followed by  $-\pi_\phi^{1,3}$  (subscripts denote pulse phase and, thus, nominal rotation axis) leads eigenstates  $|1\rangle$  and  $|3\rangle$  to assume trajectories of equal and opposite solid angle,  $\pm 2\phi$ . A similar operation,  $\pi_0^{3,4}$  followed by  $-\pi_\sigma^{3,4}$ , is applied to the nuclear spin transition, such that the total operator describing the action of these four pulses is

$$U(\phi, \sigma) = \begin{pmatrix} e^{-i\phi} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{i(\sigma+\phi)} & 0 \\ 0 & 0 & 0 & e^{-i\sigma} \end{pmatrix}$$

The value of  $\phi$  is incremented by  $\delta\phi$  on each shot of the experiment, with effective frequency  $\nu_\phi = 2\pi/\delta\phi$  (and similarly for  $\sigma$ ,  $\delta\sigma$  and  $\nu_\sigma$ ). We then map each off-diagonal element of the density matrix in turn into  $S_{x,y}I^z$  using a set of appropriate microwave and radio-frequency  $\pi$



**Figure 2 | Electron and nuclear spin phase rotations reveal the off-diagonal elements of the density matrix.** **a**, Under the application of two consecutive  $\pi^{1,3}$  pulses around different axes ( $\phi$ ), the eigenstates  $|1\rangle$  and  $|3\rangle$  undergo closed trajectories on the Bloch sphere with equal and opposite solid angles,  $\Omega = \pm 2\phi$ . Each state picks up a phase equal to half this solid angle. **b**, This  $\pi_0, -\pi_\phi$  phase gate is applied to both electron  $|1\rangle$ – $|3\rangle$  and nuclear  $|3\rangle$ – $|4\rangle$  transitions, where the two phases are varied by different increments,  $\delta\phi$  and  $\delta\sigma$ , as the experiment is repeated. Example oscillations are shown for three experiments where we generate an electron coherence,  $|1\rangle\langle 3|$ , a nuclear coherence,  $|4\rangle\langle 3|$  and a zero quantum coherence,  $|2\rangle\langle 3|$ . **c**, Fourier transforms (FT) of the oscillations with respect to increment number show peaks located at the frequencies 0.050(8), 0.031(5) and  $-0.079(8)$ , in agreement with the frequencies that were set,  $\nu_\phi = 2\pi/\delta\phi = 0.05$  and  $\nu_\sigma = 2\pi/\delta\sigma = 0.03$ .

pulses, and measure the amplitude of the Fourier component at the effective frequency corresponding to that coherence. Quadrature measurement allows us to discriminate between positive and negative frequencies. The presence of other Fourier peaks would be illustrative of pulse errors in the mapping sequence, but as seen in Fig. 2b, c, such errors are negligible even in the absence of operations such as phase cycling.

By combining our measurements of the identity component and the diagonal and off-diagonal elements of the density matrix of the electron–nuclear spin system, we obtain the following expression for  $\rho$ :

$$\begin{pmatrix} 0.382 & 0.003 + 0.000i & -0.035 - 0.039i & 0.272 \\ 0.003 - 0.000i & 0.017 & -0.000 + 0.001i & 0.001 + 0.003i \\ -0.035 + 0.039i & -0.000 - 0.001i & 0.174 & -0.055 - 0.042i \\ 0.272 & 0.001 - 0.003i & -0.055 + 0.042i & 0.427 \end{pmatrix}$$

This state has a minimum eigenvalue under the PPT test of  $-0.19(1)$  and a concurrence,  $C$ , of  $0.43(4)$ , each of which confirms the presence of finite entanglement. The results of this tomography process are shown in Fig. 3. The fidelity of the measured density matrix with respect to the target state, given that  $\alpha = 0.217$ , is  $98.2(2)\%$ , and is  $68(2)\%$  with respect to an ideal Bell state ( $\alpha = 0$ ). To obtain the uncertainty in these values, we used Monte Carlo generation of physical density matrices based on the standard error of each matrix element due to noise (Supplementary Information).

The finite entanglement shown can offer direct advantages over classical methods in applications such as quantum sensors<sup>23</sup>. To achieve higher-purity entangled states, we could use lower temperatures; for example, we would expect  $C \approx 0.99$  if these experiments were performed at  $0.8$  K. Complementary to this approach, entanglement purification could be performed using a larger Hilbert space at each node<sup>24</sup>, for example using a donor atom with a higher nuclear spin (such as bismuth, with  $I = 9/2$ ).

The electron–nuclear spin entanglement generated here could also be mapped into an entangled state between nuclear spin pairs<sup>25</sup>. By interchanging (by SWAP) the state of the electron spin with a second, coupled nucleus, for example, nuclear spin entanglement could be attained in a regime where the thermal polarization of the nuclei would be orders of magnitude too small and the direct coupling between them weak. Clusters of up to eight nuclei coupled to a single electron spin have been explored in other materials<sup>26</sup>, although the scaling of such an approach seems limited. A scalable network of entangled nuclear spins could be generated by exploiting the ability to ionize the donor and transfer the electron onto a neighbouring donor site<sup>27,28</sup>. These operations, combined with single-shot read-out of the phosphorus

donor spin<sup>29</sup> and globally controlled electron–nuclear spin entanglement such as we have demonstrated, form the basis for a cluster-state quantum computer in silicon<sup>25</sup>.

## METHODS SUMMARY

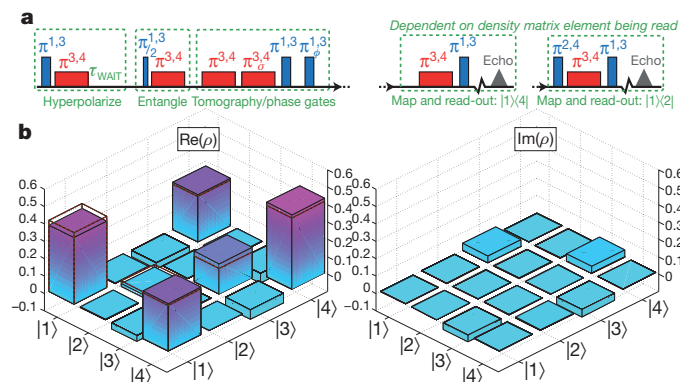
Si:P consists of an electron spin,  $S = 1/2$  ( $g = 1.9987$ ), coupled to the nuclear spin,  $I = 1/2$ , of  $^{31}\text{P}$  through an isotropic hyperfine coupling of  $a = 4.19$  mT. The W-band electron spin resonance signal comprises two lines (one for each nuclear spin projection,  $M_I = \pm 1/2$ ). Our experiments were performed on the low-field line of the electron spin resonance doublet, corresponding to  $M_I = 1/2$ . At  $2.9$  K and  $3.36$  T, the electron and nuclear spin relaxation times were measured to be approximately  $0.6$  s and  $100$  s, respectively.

The sample consists of a  $^{28}\text{Si}$ -enriched single crystal about  $0.5$  mm in diameter with a residual  $^{29}\text{Si}$  concentration of order  $70$  p.p.m., produced by decomposing isotopically enriched silane in a recirculating reactor to produce poly-silicon rods, followed by floating-zone crystallization. Phosphorus doping of  $\sim 10^{14}$  cm $^{-3}$  was achieved by adding dilute  $\text{PH}_3$  gas to the ambient argon during the final floating-zone single-crystal growth. Further information on the sample growth has been reported elsewhere<sup>30</sup>.

Pulsed electron spin resonance experiments were performed using a W-band (94-GHz) Bruker ELEXSYS 680 spectrometer, modified to allow microwave phase control and equipped with a 6-T superconducting magnet and a low-temperature helium-flow cryostat (Oxford CF935). The cryostat was pumped to achieve a temperature of  $2.88$  K (internal thermocouple) consistent with the spin temperature measurement (see text). Typical pulse times were  $56$  ns for a microwave  $\pi$  pulse and  $100$   $\mu$ s for a radio-frequency  $\pi$  pulse. To achieve arbitrary phase control, we generated radio-frequency pulses using a Rohde and Schwarz AFQ100B together with an Amplifier Research 500 W amplifier.

Received 30 September; accepted 23 November 2010.

Published online 19 January 2011.



**Figure 3 | Measuring an entangled density matrix.** **a**, The full pulse sequence used to prepare, entangle and measure the two-spin state. The final read-out stage was changed according to the density matrix element being measured: examples are shown for the  $|1\rangle|2\rangle$  and  $|1\rangle|4\rangle$  states. **b**, The obtained density matrix is shown as solid bars, and the dashed outline (zero where not shown) shows that of an ideal state given  $\alpha = 0.217$ . The fidelity of the ideal state with the measured density matrix is  $98\%$ .

- Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Jozsa, R. & Linden, N. On the role of entanglement in quantum-computational speed-up. *Proc. R. Soc. Lond. A* **459**, 2011–2032 (2003).
- Curtis, M., Lewenstein, M. & Lütkenhaus, N. Entanglement as a precondition for secure quantum key distribution. *Phys. Rev. Lett.* **92**, 217903 (2004).
- Vandersypen, L. *et al.* Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
- Negrevergne, C. *et al.* Benchmarking quantum control methods on a 12-qubit system. *Phys. Rev. Lett.* **96**, 170501 (2006).
- Knill, E., Chuang, I. & Laflamme, R. Effective pure states for bulk quantum computation. *Phys. Rev. A* **57**, 3348–3363 (1998).
- Horodecki, M., Horodecki, P. & Horodecki, R. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A* **223**, 1–8 (1996).
- Peres, A. Separability criterion for density matrices. *Phys. Rev. Lett.* **77**, 1413–1415 (1996).
- Mehring, M., Mende, J. & Scherer, W. Entanglement between an electron and a nuclear spin  $1/2$ . *Phys. Rev. Lett.* **90**, 153001 (2003).
- Anwar, M. *et al.* Preparing high purity initial states for nuclear magnetic resonance quantum computing. *Phys. Rev. Lett.* **93**, 040501 (2004).
- Morton, J. J. L. *et al.* Solid-state quantum memory using the  $^{31}\text{P}$  nuclear spin. *Nature* **455**, 1085–1088 (2008).
- Tyryshkin, A. M. & Lyon, S. A. Data presented at the Silicon Qubit Workshop, 23–24 August (Albuquerque, sponsored by Lawrence Berkeley National Laboratory and Sandia National Laboratory, 2010).
- Tyryshkin, A. M. *et al.* Coherence of spin qubits in silicon. *J. Phys. Condens. Matter* **18**, S783–S794 (2006).
- Morton, J. J. L. *et al.* High fidelity single qubit operations using pulsed electron paramagnetic resonance. *Phys. Rev. Lett.* **95**, 200501 (2005).
- Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
- Wollan, D. S. Dynamic nuclear polarization with an inhomogeneously broadened ESR line. II. Experiment. *Phys. Rev. B* **13**, 3686–3696 (1976).
- Hayashi, H., Itahashi, T., Itoh, K. M., Vlasenko, L. S. & Vlasenko, M. P. Dynamic nuclear polarization of  $^{29}\text{Si}$  nuclei in isotopically controlled phosphorus doped silicon. *Phys. Rev. B* **80**, 045201 (2009).
- Schulman, L., Mor, T. & Weinstein, Y. Physical limits of heat-bath algorithmic cooling. *Phys. Rev. Lett.* **94**, 120501 (2005).
- Wei, T. *et al.* Maximal entanglement versus entropy for mixed quantum states. *Phys. Rev. A* **67**, 022110 (2003).
- Verstraete, F., Audenaert, K. & De Moor, B. Maximally entangled mixed states of two qubits. *Phys. Rev. A* **64**, 012316 (2001).
- Aharonov, Y. & Anandan, J. Phase change during a cyclic quantum evolution. *Phys. Rev. Lett.* **58**, 1593–1596 (1987).
- Suter, D., Mueller, K. T. & Pines, A. Study of the Aharonov–Anandan quantum phase by NMR interferometry. *Phys. Rev. Lett.* **60**, 1218–1220 (1988).

23. Simmons, S., Jones, J. A., Karlen, S. D., Ardavan, A. & Morton, J. J. L. Magnetic field sensors using 13-spin cat states. *Phys. Rev. A* **82**, 022330 (2010).
24. Campbell, E. T. Distributed quantum-information processing with minimal local resources. *Phys. Rev. A* **76**, 040302 (2007).
25. Morton, J. J. L. A silicon-based cluster state quantum computer. Preprint at (<http://128.84.158.119/abs/0905.4008v1>) (2009).
26. Mehring, M. & Mende, J. Spin-bus concept of spin quantum computing. *Phys. Rev. A* **73**, 052303 (2006).
27. Skinner, A., Davenport, M. & Kane, B. Hydrogenic spin quantum computing in silicon: a digital approach. *Phys. Rev. Lett.* **90**, 087901 (2003).
28. Andresen, S. *et al.* Charge state control and relaxation in an atomically doped silicon device. *Nano Lett.* **7**, 2000–2003 (2007).
29. Morello, A. *et al.* Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).
30. Becker, P., Pohl, H.-J., Riemann, H. & Abrosimov, N. Enrichment of silicon for a better kilogram. *Phys. Status Solidi. A* **207**, 49–66 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Fitzsimons, S. Benjamin, A. Ardavan, A. Briggs and B. Lovett for discussions, and P. Höfer and Bruker BioSpin for support with instrumentation. Three-dimensional images were created using POV-RAY open-source software. We thank EPSRC for supporting work at Oxford through CAESR (EP/D048559/1) and the Oxford–Keio collaboration through the JST-EPSRC SIC programme (EP/H025952/1). Work at Keio has been supported by Grants-in-aid for Scientific Research by MEXT, FIRST by JSPS, Nanoquine and Keio GCOE. S.S. is supported by the Clarendon Fund, J.J.L.M. is supported by St John's College, Oxford, and the Royal Society.

**Author Contributions** S.S., R.M.B. and J.J.L.M. designed and performed the experiments and wrote the paper. H.R., N.V.A., P.B. and H.-J.P. grew the  $^{28}\text{Si}$  crystal. K.M.I. and M.L.W.T. analysed and prepared the sample and discussed the experiments, results and manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.J.L.M. ([john.morton@materials.ox.ac.uk](mailto:john.morton@materials.ox.ac.uk)).



## Femtosecond X-ray protein nanocrystallography

Henry N. Chapman<sup>1,2</sup>, Petra Fromme<sup>3</sup>, Anton Barty<sup>1</sup>, Thomas A. White<sup>1</sup>, Richard A. Kirian<sup>4</sup>, Andrew Aquila<sup>1</sup>, Mark S. Hunter<sup>3</sup>, Joachim Schulz<sup>1</sup>, Daniel P. DePonte<sup>1</sup>, Uwe Weierstall<sup>4</sup>, R. Bruce Doak<sup>4</sup>, Filipe R. N. C. Maia<sup>5</sup>, Andrew V. Martin<sup>1</sup>, Ilme Schlichting<sup>6,7</sup>, Lukas Lomb<sup>7</sup>, Nicola Coppola<sup>1†</sup>, Robert L. Shoeman<sup>7</sup>, Sascha W. Epp<sup>6,8</sup>, Robert Hartmann<sup>9</sup>, Daniel Rolles<sup>6,7</sup>, Artem Rudenko<sup>6,8</sup>, Lutz Foucar<sup>6,7</sup>, Nils Kimmel<sup>10</sup>, Georg Weidenspointner<sup>11,10</sup>, Peter Holl<sup>9</sup>, Mengning Liang<sup>1</sup>, Miriam Barthelmess<sup>12</sup>, Carl Caleman<sup>1</sup>, Sébastien Boutet<sup>13</sup>, Michael J. Bogan<sup>14</sup>, Jacek Krzywinski<sup>13</sup>, Christoph Bostedt<sup>13</sup>, Saša Bajt<sup>12</sup>, Lars Gumprecht<sup>1</sup>, Benedikt Rudek<sup>6,8</sup>, Benjamin Erk<sup>6,8</sup>, Carlo Schmidt<sup>6,8</sup>, André Hömke<sup>6,8</sup>, Christian Reich<sup>9</sup>, Daniel Pietschner<sup>10</sup>, Lothar Strüder<sup>6,10</sup>, Günter Hauser<sup>10</sup>, Hubert Gorke<sup>15</sup>, Joachim Ullrich<sup>6,8</sup>, Sven Herrmann<sup>10</sup>, Gerhard Schaller<sup>10</sup>, Florian Schopper<sup>10</sup>, Heike Soltau<sup>9</sup>, Kai-Uwe Kühnel<sup>8</sup>, Marc Messerschmidt<sup>13</sup>, John D. Bozek<sup>13</sup>, Stefan P. Hau-Riege<sup>16</sup>, Matthias Frank<sup>16</sup>, Christina Y. Hampton<sup>14</sup>, Raymond G. Sierra<sup>14</sup>, Dmitri Starodub<sup>14</sup>, Garth J. Williams<sup>13</sup>, Janos Hajdu<sup>5</sup>, Nicusor Timneanu<sup>5</sup>, M. Marvin Seibert<sup>5†</sup>, Jakob Andreasson<sup>5</sup>, Andrea Rocker<sup>5</sup>, Olof Jönsson<sup>5</sup>, Martin Svenda<sup>5</sup>, Stephan Stern<sup>1</sup>, Karol Nass<sup>2</sup>, Robert Andrichke<sup>10</sup>, Claus-Dieter Schröter<sup>8</sup>, Faton Krasniqi<sup>6,7</sup>, Mario Bott<sup>7</sup>, Kevin E. Schmidt<sup>4</sup>, Xiaoyu Wang<sup>4</sup>, Ingo Grotjohann<sup>3</sup>, James M. Holton<sup>17</sup>, Thomas R. M. Barends<sup>7</sup>, Richard Neutze<sup>18</sup>, Stefano Marchesini<sup>17</sup>, Raimund Fromme<sup>3</sup>, Sebastian Schorb<sup>19</sup>, Daniela Rupp<sup>19</sup>, Marcus Adolph<sup>19</sup>, Tais Gorkhover<sup>19</sup>, Inger Andersson<sup>20</sup>, Helmut Hirsemann<sup>12</sup>, Guillaume Potdevin<sup>12</sup>, Heinz Graafsma<sup>12</sup>, Björn Nilsson<sup>12</sup> & John C. H. Spence<sup>4</sup>

X-ray crystallography provides the vast majority of macromolecular structures, but the success of the method relies on growing crystals of sufficient size. In conventional measurements, the necessary increase in X-ray dose to record data from crystals that are too small leads to extensive damage before a diffraction signal can be recorded<sup>1–3</sup>. It is particularly challenging to obtain large, well-diffracting crystals of membrane proteins, for which fewer than 300 unique structures have been determined despite their importance in all living cells. Here we present a method for structure determination where single-crystal X-ray diffraction ‘snapshots’ are collected from a fully hydrated stream of nanocrystals using femtosecond pulses from a hard-X-ray free-electron laser, the Linac Coherent Light Source<sup>4</sup>. We prove this concept with nanocrystals of photosystem I, one of the largest membrane protein complexes<sup>5</sup>. More than 3,000,000 diffraction patterns were collected in this study, and a three-dimensional data set was assembled from individual photosystem I nanocrystals (~200 nm to 2 µm in size). We mitigate the problem of radiation damage in crystallography by using pulses briefer than the timescale of most damage processes<sup>6</sup>. This offers a new approach to structure determination of macromolecules that do not yield crystals of sufficient size for studies using conventional radiation sources or are particularly sensitive to radiation damage.

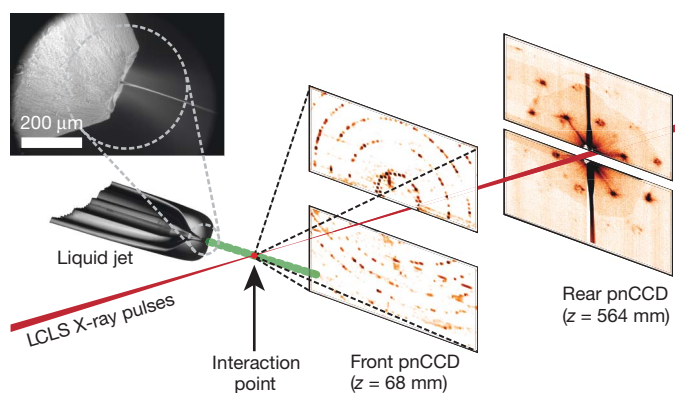
Radiation damage has always limited resolution in biological imaging using electrons or X-rays<sup>2</sup>. With the recent invention of the femtosecond X-ray laser, an opportunity has arisen to break the nexus between radiation dose and spatial resolution. It has been proposed that femtosecond X-ray pulses can be used to outrun even the fastest damage processes by using single pulses so brief that they terminate before the manifestation of damage to the sample<sup>6</sup>. Experiments at the FLASH free-electron laser (FEL), Germany, confirmed the feasibility of ‘diffraction before destruction’ at resolution lengths down to 60 Å on test samples fixed on silicon nitride membranes<sup>7</sup>. It was predicted that

the irradiance (or power density) of focused pulses from a hard-X-ray FEL such as the Linac Coherent Light Source (LCLS), USA, would be sufficient to produce diffraction patterns at near-atomic resolution<sup>6</sup>.

We demonstrate here that this notion of diffraction before destruction operates at subnanometre resolution, using the membrane protein photosystem I as a model system, and establish an approach to structure determination based on X-ray diffraction data from a stream of nanocrystals<sup>6,8</sup>. Membrane proteins have a central role in the functioning of cells and viruses, yet our knowledge of the structure and dynamics responsible for their functioning remains limited. Photosystem I is a large membrane protein complex (1-MDa molecular mass, 36 proteins, 381 cofactors) that acts as a biosolar energy converter in the process of oxygenic photosynthesis. Its crystals display the symmetry of space group *P*6<sub>3</sub>, with unit-cell parameters *a* = *b* = 281 Å and *c* = 165 Å, and consist of 78% solvent by volume. We show that diffraction data can be recorded from these fragile protein nanocrystals before destruction occurs. Furthermore, we demonstrate that structure factors can be extracted from the ‘partial’ reflections of tens of thousands of single-crystal diffraction snapshots, showing that interpretable high-quality, three-dimensional (3D) structure factor data can be obtained from a suspension of submicrometre crystals.

Our experimental set-up (Fig. 1 and Methods) records single-crystal diffraction data from a stream of crystals carried in a 4-µm-diameter, continuous liquid water jet<sup>9</sup> that flows across the focused LCLS X-ray beam in vacuum at 10 µl min<sup>−1</sup>. In contrast to cryo-electron microscopy<sup>10,11</sup> or standard crystallography on microcrystals<sup>3</sup>, which require cryogenic cooling, these data were collected on fully hydrated, 3D nanocrystals. The crystal located in the interaction region when an X-ray pulse arrives gives rise to a diffraction pattern that is detected on a set of two low-noise, X-ray p–n junction charge-coupled device (pnCCD) modules<sup>12</sup> and read out before the arrival of the next pulse at the FEL repetition rate of 30 Hz, or 1,800 patterns per minute. The

<sup>1</sup>Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. <sup>2</sup>University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. <sup>3</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA. <sup>4</sup>Department of Physics, Arizona State University, Tempe, Arizona 85287, USA. <sup>5</sup>Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden. <sup>6</sup>Max Planck Advanced Study Group, Center for Free-Electron Laser Science, Notkestrasse 85, 22607 Hamburg, Germany. <sup>7</sup>Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. <sup>8</sup>Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany. <sup>9</sup>PNSensor GmbH, Otto-Hahn-Ring 6, 81739 München, Germany. <sup>10</sup>Max-Planck-Institut Halbleiterlabor, Otto-Hahn-Ring 6, 81739 München, Germany. <sup>11</sup>Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstrasse, 85741 Garching, Germany. <sup>12</sup>Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. <sup>13</sup>LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>14</sup>PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>15</sup>Forschungszentrum Jülich, Institut ZEL, 52425 Jülich, Germany. <sup>16</sup>Lawrence Livermore National Laboratory, 7000 East Avenue, Mail Stop L-211, Livermore, California 94551, USA. <sup>17</sup>Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>18</sup>Department of Chemistry, Biochemistry and Biophysics, University of Gothenburg, SE-405 30 Gothenburg, Sweden. <sup>19</sup>Institut für Optik und Atomare Physik, Technische Universität Berlin, Hardenbergstrasse 36, 10623 Berlin, Germany. <sup>20</sup>Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden. †Present addresses: European XFEL GmbH, Notkestrasse 85, 22607 Hamburg, Germany (N.C.); LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA (M.M.S.).



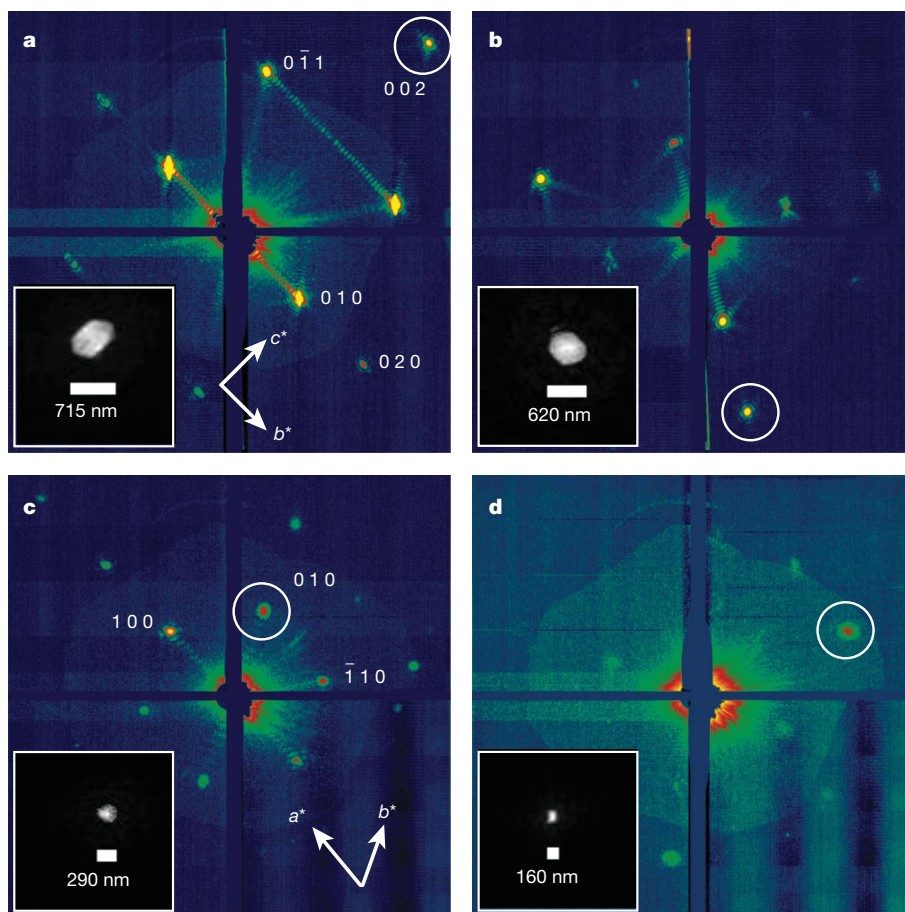
**Figure 1 | Femtosecond nanocrystallography.** Nanocrystals flow in their buffer solution in a gas-focused, 4- $\mu\text{m}$ -diameter jet at a velocity of  $10\text{ m s}^{-1}$  perpendicular to the pulsed X-ray FEL beam that is focused on the jet. Inset, environmental scanning electron micrograph of the nozzle, flowing jet and focusing gas<sup>30</sup>. Two pairs of high-frame-rate pnCCD detectors<sup>12</sup> record low- and high-angle diffraction from single X-ray FEL pulses, at the FEL repetition rate of 30 Hz. Crystals arrive at random times and orientations in the beam, and the probability of hitting one is proportional to the crystal concentration.

photon energy of the X-ray pulses was 1.8 keV (6.9- $\text{\AA}$  wavelength), with more than  $10^{12}$  photons per pulse at the sample and pulse durations of 10, 70, and 200 fs (ref. 13). An X-ray fluence of  $900\text{ J cm}^{-2}$  was achieved by focusing the FEL beam to a full-width at half-maximum of 7  $\mu\text{m}$ , corresponding to a sample dose of up to 700 MGy per pulse (calculated

using the program RADDOSE<sup>14</sup>) and a peak power density in excess of  $10^{16}\text{ W cm}^{-2}$  at 70-fs duration. In contrast, the typical tolerable dose in conventional X-ray experiments is only about 30 MGy (ref. 1). A single LCLS X-ray pulse destroys any solid material placed in this focus, but the stream replenishes the vaporized sample before the next pulse.

The front detector module, located close to the interaction region, recorded high-angle diffraction to a resolution of 8.5  $\text{\AA}$ , whereas the rear module intersected diffraction at resolutions in the range of 4,000 to 100  $\text{\AA}$ . We observed diffraction from crystals smaller than ten unit cells on a side, as determined by examining the data recorded on the rear pnCCDs (Fig. 2). A crystal with a side length of  $N$  unit cells gives rise to diffraction features that are finer by a factor of  $1/N$  than the Bragg spacing (that is, with  $N - 2$  fringes between neighbouring Bragg peaks), providing a simple way to determine the projected size of the nanocrystal. Images of crystal shapes obtained using an iterative phase retrieval method<sup>15,16</sup> are shown in Fig. 2. The 3D Fourier transform of the crystal shape is repeated on every reciprocal lattice point. However, the diffraction condition for lattice points is usually not exactly satisfied, so each recorded Bragg spot represents a particular ‘slice’ of the Ewald sphere through the shape transform, giving a variety of Bragg spot profiles in a pattern; these are apparent in Fig. 2. The sum of counts in each Bragg spot underestimates the underlying structure factor square modulus, representing a partial reflection.

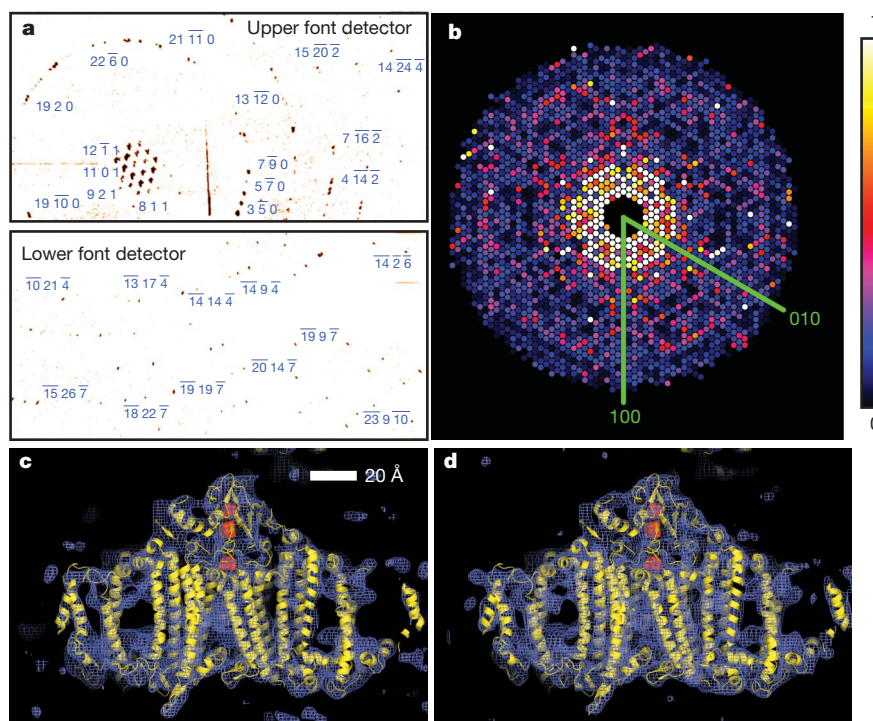
Figure 3a shows strong single-crystal diffraction to the highest angles of the front detector. The nanocrystal shape transform is also apparent in many patterns at the high angles detected by the front detector, giving significant measured intensities between Bragg peaks as is noticeable in Supplementary Fig. 3a. These mid-Bragg intensities



**Figure 2 | Coherent crystal diffraction.** Low-angle diffraction patterns recorded on the rear pnCCDs, revealing coherent diffraction from the structure of the photosystem I nanocrystals, shown using a logarithmic, false-colour scale. The Miller indices of the peaks in **a** were identified from the

corresponding high-angle pattern. In **c** we count seven fringes in the  $b^*$  direction, corresponding to nine unit cells, or 250 nm. Insets, real-space images of the nanocrystal, determined by phase retrieval (using the Shrinkwrap algorithm<sup>15</sup>) of the circled coherent Bragg shape transform.





**Figure 3 | Diffraction intensities and electron density of photosystem I.**

**a**, Diffraction pattern recorded on the front pnCCDs with a single 70-fs pulse after background subtraction and correction of saturated pixels. Some peaks are labelled with their Miller indices. The resolution in the lower detector corner is 8.5 Å. **b**, Precession-style pattern of the [001] zone for photosystem I, obtained from merging femtosecond nanocrystal data from over 15,000 nanocrystal

patterns, displayed on the linear colour scale shown on the right. **c**, **d**, Region of the  $2mF_o - DF_c$  electron density map at  $1.0\sigma$  (purple mesh), calculated from the 70-fs data (**c**) and from conventional synchrotron data truncated at a resolution of 8.5 Å and collected at a temperature of 100 K (**d**) (Methods). The refined model is depicted in yellow.

oversample the molecular transform, providing a potential route to phasing of the pattern<sup>17,18</sup>.

In conventional crystallography, the 'full' Bragg reflection is determined to high precision, for example by integrating counts as the crystal is rotated such that these reflections pass through the diffraction condition. By indexing individual patterns and then summing counts in all partial reflections for each index, we performed a Monte Carlo integration over the reciprocal-space volume of the Bragg reflection and the distribution of crystal shapes and orientations and variations in the X-ray pulse fluence. The result of this procedure converges to the square of the structure factor moduli<sup>18</sup>. We found that over 13% of diffraction patterns with ten or more spots could be consistently indexed using the programs MOSFLM<sup>19</sup> and DirAx<sup>20</sup> (Methods). Merged intensities at 70-fs pulse duration are presented as a precession-style image of the [001]-zone axis in Fig. 3b (see also Supplementary Figs 3 and 4). We tested the reliability of this approach by comparing the LCLS merged data with data collected at 100 K with 12.4-keV synchrotron radiation from a single crystal of photosystem I cryopreserved in 2 M sucrose. These data sets show good agreement, with a difference metric,  $R_{iso}$ , of 22.1% computed over the entire resolution range and of less than 13% in the middle resolution shells; see Supplementary Table 1 for detailed statistics.

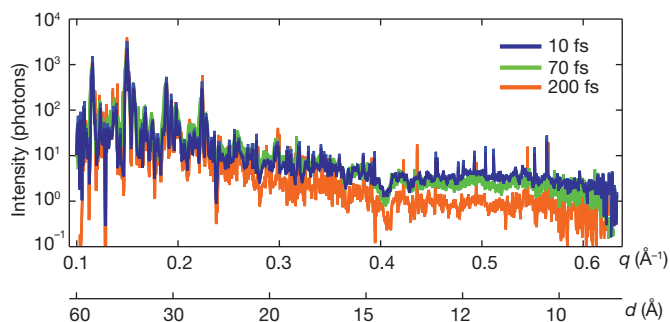
To complete our proof of principle, we conducted a rigid-body refinement of the published photosystem I structure (Protein Data Bank ID, 1JB0) against the nanocrystal structure factors, yielding  $R/R_{free} = 0.25/0.23$ . A representative region of the  $2mF_o - DF_c$  electron density map at 8.5 Å (Methods) from the LCLS data set is shown in Fig. 3c. This map shows the details expected at this resolution, including transmembrane helices, membrane extrinsic features and some loop structures. For comparison, the electron density refined from the 12.4-keV, single-crystal data set truncated to a resolution of 8.5 Å is given in Fig. 3d.

The dose of 700 MGy corresponds to a K-shell photoabsorption of 3% of all carbon atoms in the protein. This energy is subsequently

released by photoionization and Auger decay, followed by a cascade of lower-energy electrons caused by secondary ionizations, taking place on the 10–100-fs timescale<sup>21</sup>. Using a model of the plasma dynamics<sup>22,23</sup>, we calculated that by the end of a 100-fs pulse each atom of the crystal was ionized once, on average, and that motion of nuclei had begun. This is expected to give rise to a decrease in Bragg amplitudes, similar to an increase in a Debye–Waller temperature factor<sup>24</sup>. We studied the effects of the initial ionization damage on the diffraction of photosystem I nanocrystals by collecting a series of data sets at pulse durations of 10, 70 and 200 fs. The 10-fs pulses were produced with lower pulse energy: ~10% of the total number of photons of the longer pulses<sup>13</sup>, or a 70-MGy dose. Plots of the scattering strength of the crystals versus resolution, generated by selecting and summing Bragg spots from more than 66,000 patterns for each of the three pulse durations measured, are shown in Fig. 4. The 10- and 70-fs traces are very similar, indicating that these pulses are short enough to overcome radiation damage at the observed resolution, 8.5 Å. For 200-fs pulses, there is a decrease in scattering strength at resolutions beyond 25 Å, indicating disordering on this longer timescale. The highest-resolution Bragg peaks for the 200-fs pulses were not broadened or shifted relative to the short-duration data sets, which indicates there was no strain or expansion of the lattice, respectively.

Our next step is to improve resolution by using shorter-wavelength X-rays. Resolution may ultimately be limited by X-ray pulse fluence, the ultrafast radiation damage and the intrinsic disorder within the nanocrystals themselves. Recent experiments<sup>21</sup> at LCLS indicate a brief saturation of the X-ray photoabsorption of atoms in a tightly focused pulse, resulting in a decrease in photoionization damage on a 20-fs timescale without a reduction in the scattering cross-sections that give rise to the diffraction pattern<sup>22</sup>. Planned beamlines at LCLS aim to achieve up to a  $10^5$ -fold increase in pulse irradiance by tighter focusing, allowing data collection with low-fluence, 10-fs pulses or pulses of even shorter duration<sup>25</sup>. This provides a route to further reducing radiation damage and may allow measurements on even smaller nanocrystals,





**Figure 4 | Pulse-duration dependence of diffraction intensities.** Plot of the integrated Bragg intensities of photosystem I nanocrystal diffraction as a function of photon momentum transfer,  $q = (4\pi/\lambda)\sin(\theta) = 2\pi/d$  (wavelength,  $\lambda$ ; scattering angle,  $2\theta$ ; resolution,  $d$ ) for pulse durations of 10, 70 and 200 fs. Averages were obtained by isolating Bragg spots from 97,883, 805,311 and 66,063 patterns, respectively, normalized to pulse fluence. The error in each plot is indicated by the thickness of the line. The decrease in irradiance for 200-fs pulses and  $d < 25$  Å indicates radiation damage for these long pulses, which is not apparent for 70-fs pulses and shorter.

down to a single unit cell<sup>6</sup> (that is, a single molecule). As this limit is approached, the ordering of the nanocrystals will become increasingly irrelevant, as each crystal may be treated as a single object and the ‘disorder’ that conventionally leads to reduced resolution will simply manifest itself as shot-to-shot variability, providing information about not just the average structure but also the range of dynamically accessible conformations.

Data are collected on fully hydrated nanocrystals without cryogenic cooling. We expect that the results presented here will open new avenues for crystallography using X-ray laser pulses that are so short that only negligible X-ray-induced radiation damage occurs during data collection. Significant improvements in sample utilization are expected by exploiting higher X-ray repetition rates or by slowing the liquid flow. For example, the generation, using inkjet technologies, of liquid droplets at a rate that matches the LCLS X-ray pulses would dramatically decrease the total required sample volume by a factor of 25,000, meaning that less than 0.4 µl of nanocrystal suspension would be needed in our particular case, of photosystem I. Further efficiency gains would result from indexing and merging a greater proportion of patterns into the 3D data set, which may be achieved by applying methods for merging continuous diffraction patterns of single molecules<sup>26,27</sup> or by using ‘post-refinement’<sup>28</sup> to obtain accurate structure factor estimates from fewer diffraction patterns. These methods will also remove the twinning ambiguity that exists in our current indexing scheme. Our method also has potential application to the study of chemical reactions, such as the processes in photosynthesis or enzymatic reactions.

## METHODS SUMMARY

We made our measurements using the CFEL-ASG Multi-Purpose (CAMP) instrument<sup>12</sup> on the Atomic, Molecular and Optical Science beamline<sup>29</sup> at the LCLS<sup>4</sup>. Diffraction data were recorded at the LCLS repetition rate of 30 Hz with a set of two movable, high-frame-rate, low-noise, X-ray pnCCD detector units<sup>12</sup>. The front detector, located 68 mm from the jet, accepts scattering angles up to 47.9°, corresponding to a resolution of 8.5 Å at a wavelength of 6.9 Å. The rear unit was located 564 mm from the jet to record finer sampling of the diffraction pattern at low angles.

The liquid jet was emitted from a capillary with an inner diameter of 40 µm and focused by a coaxial flow of gas to a diameter of about 4 µm (ref. 9), flowing at 10 µl min<sup>-1</sup>. The low jet diameter constrains the crystals to pass through the most intense part of the focused X-ray beam. Clogging of nanocrystals in the capillary is avoided, and the coaxial gas sheath prevents freezing of the liquid in the vacuum environment. A micropore filter in the fluid delivery line was used to restrict the size of the photosystem I nanocrystals to less than 2 µm. The suspension was diluted to observe a crystal ‘hit rate’ of 20% (Supplementary Fig. 2) to reduce the occurrence of double hits. The concentration of observed crystals was therefore 0.2 per illuminated volume of  $4 \times 4 \times 13$  µm<sup>3</sup>, or about 10<sup>9</sup> crystals per millilitre.

The overall protein concentration after dilution of the suspension was 1 mg ml<sup>-1</sup> (1 µM of the photosystem I trimer), and a complete set of structure factors was obtained from 1,850,000 X-ray pulses.

Diffraction peaks from the 70-fs data were identified, indexed and combined into a set of 3D structure factors comprising 3,379 unique reflections from 2,424,394 spots. Statistics of the merged data are given in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 24 July; accepted 9 December 2010.

- Owen, R. L., Rudino-Pinera, E. & Garman, E. F. Experimental determination of the radiation dose limit for cryocooled protein crystals. *Proc. Natl Acad. Sci. USA* **103**, 4912–4917 (2006).
- Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28**, 171–193 (1995).
- Riek, C. Recent developments in microdiffraction on protein crystals. *J. Synchr. Radiat.* **11**, 4–6 (2004).
- Emma, P. *et al.* First lasing and operation of an ångström-wavelength free-electron laser. *Nature Photon.* **4**, 641–647 (2010).
- Jordan, P. *et al.* Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature* **411**, 909–917 (2001).
- Neutze, R., Wout, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
- Chapman, H. N. *et al.* Femtosecond time-delay X-ray holography. *Nature* **448**, 676–679 (2007).
- Spence, J. C. H. & Doak, R. B. Single molecule diffraction. *Phys. Rev. Lett.* **92**, 198102 (2004).
- DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D* **41**, 195505 (2008).
- Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* **213**, 899–929 (1990).
- Wang, D. N. & Kühlbrandt, W. High-resolution electron crystallography of light-harvesting chlorophyll a/b-protein complex in three different media. *J. Mol. Biol.* **217**, 691–699 (1991).
- Strüder, L. *et al.* Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources. *Nucl. Instrum. Methods Phys. Res. A* **614**, 483–496 (2010).
- Ding, Y. *et al.* Measurements and simulations of ultralow emittance and ultrashort electron beams in the Linac Coherent Light Source. *Phys. Rev. Lett.* **102**, 254801 (2009).
- Paithankar, K. S., Owen, R. L. & Garman, E. F. Absorbed dose calculations for macromolecular crystals: improvements to RADDOSE. *J. Synchr. Radiat.* **16**, 152–162 (2009).
- Marchesini, S. *et al.* X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B* **68**, 140101 (2003).
- Robinson, I. K. & Harder, R. Coherent X-ray diffraction imaging of strain at the nanoscale. *Nature Mater.* **8**, 291–298 (2009).
- Sayre, D. Some implications of a theorem due to Shannon. *Acta Crystallogr.* **5**, 843 (1952).
- Kirian, R. *et al.* Femtosecond protein nanocrystallography—data analysis methods. *Opt. Express* **18**, 5713–5723 (2010).
- Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62**, 48–57 (2006).
- Duisenberg, A. J. M. Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Cryst.* **25**, 92–96 (1992).
- Young, L. *et al.* Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature* **466**, 56–61 (2010).
- Hau-Riege, S. P., London, R. A. & Szoke, A. Dynamics of biological molecules irradiated by short X-ray pulses. *Phys. Rev. E* **69**, 051906 (2004).
- Bergh, M., Hultdt, G., Timneanu, N., Maia, F. R. N. C. & Hajdu, J. Feasibility of imaging living cells at subnanometer resolutions by ultrafast X-ray diffraction. *Q. Rev. Biophys.* **41**, 181–204 (2008).
- Willis, B. & Pryor, A. *Thermal Vibrations in Crystallography* 92 (Cambridge Univ. Press, 1975).
- Emma, P. *et al.* Femtosecond and subfemtosecond X-ray pulses from a self-amplified spontaneous-emission based free-electron laser. *Phys. Rev. Lett.* **92**, 074801 (2004).
- Loh, N.-T. D. & Elser, V. Reconstruction algorithm for single-particle diffraction imaging experiments. *Phys. Rev. E* **80**, 026705 (2009).
- Fung, R., Shneerson, V., Saldin, D. K. & Ourmazd, A. Structure from fleeting illumination of faint spinning objects in flight. *Nature Phys.* **5**, 64–67 (2008).
- Rossmann, M. G., Leslie, A. G., Sherin, S. A. & Tsukihara, T. Processing and post-refinement of oscillation camera data. *J. Appl. Cryst.* **12**, 570–581 (1979).
- Bozek, J. D. AMO instrumentation for the LCLS X-ray FEL. *Eur. Phys. J. Spec. Top.* **169**, 129–132 (2009).
- DePonte, D. P. *et al.* SEM imaging of liquid jets. *Micron* **40**, 507–509 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Experiments were carried out at the Linac Coherent Light Source and the Advanced Light Source, both National User Facilities operated respectively by Stanford University and the University of California on behalf of the US Department of

Energy (DOE), Office of Basic Energy Sciences. We acknowledge support from the DOE through the PULSE Institute at the SLAC National Accelerator Laboratory; the Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344; the Center for Bio-Inspired Solar Fuel Production, an Energy Frontier Research Center funded by the DOE, Office of Basic Energy Sciences (award DE-SC0001016); the Hamburg Ministry of Science and Research and the Joachim Herz Stiftung, as part of the Hamburg Initiative for Excellence in Research (LEXI); the Hamburg School for Structure and Dynamics; the Max Planck Society, for funding the development and operation of the CAMP instrument within the ASG at CFEL; the US National Science Foundation (awards 0417142 and MCB-1021557); the US National Institutes of Health (awards 1R01GM095583-01 (ROADMAP) and 1U54GM094625-01 (PSI:BiologY)); the Swedish Research Council; the Swedish Foundation for International Cooperation in Research and Higher Education; Stiftelsen Olle Engkvist Byggmästare; the DFG Cluster of Excellence at the Munich Centre for Advanced Photonics; and the CBST at the University of California under cooperative agreement no. PHY 0120999. We acknowledge discussions with M. Rossmann, E. Snell, R. Stroud and A. Brunger, thank B. Hedman, E. Gullikson, F. Filsinger, A. Berg, H. Mahn and C. Kaiser for technical help and thank the staff of the LCLS for their support in carrying out these experiments.

**Author Contributions** H.N.C. and J.C.H.S. conceived the experiment, which was designed with P.F., A.B., R.A.K., J.S., D.P.D., U.W., R.B.D., S. Boutet, M.J.B., D.S., I.S., S.M. and J.H. The CAMP instrument was the responsibility of S.W.E., R.H., D. Rolles, A. Rudenko, C.S., L.F., N.K., P.H., B.R., B.E., A.H., Ch.R., D.P., G.W., L.S., G.H., H. Gorker, J.U., I.S.,

S.H., G.S., F.S., H.S., K.-U.K., R.A., C.-D.S., F.K., M. Bott, S. Schorb, D. Rupp, M.A., T.G., H.H., L.G., G.P., H. Graafsma and B.N., who designed and set up the instrument and/or developed and operated the pnCCD detectors. C.B., J.D.B. and M.M. set up and aligned the beamline. P.F., M.S.H. and I.G. prepared samples; R.B.D., D.P.D., U.W., J.C.H.S., P.F., L.L. and R.L.S. developed and operated the sample delivery system; H.N.C., A.B., A.A., J.S., D.P.D., U.W., R.B.D., S. Bajt, M.J.B., L.G., J.H., M.M.S., N.T., J.A., S. Stern and J.C.H.S. developed diffraction instrumentation; and M. Barthelmess, M.L., A.B. and K.N. designed and/or fabricated calibration samples. J.K., S.P.H.-R., A.B., H.N.C., J.S. and A.V.M. characterized the focus. H.N.C., J.C.H.S., P.F., A.B., T.A.W., R.A.K., A.A., J.S., D.P.D., U.W., R.B.D., I.S., N.C., R.L.S., M.S.H., L.L., M. Bott, S.W.E., R.H., D. Rolles, A. Rudenko, M.L., C.B., J.U., L.F., J.D.B., M.M., M.F., C.Y.H., R.G.S., G.J.W., A. Rocker, M.S., O.J., I.A. and J.H. carried out the experiment. A.B., T.A.W., R.A.K., A.A., F.R.N.C.M., A.V.M., L.L., T.R.M.B., N.C., L.F., N.K., R.N., G.W., P.H., C.C., J.M.H., I.S., J.H., H.N.C. and J.C.H.S. analysed the data. A.V.M. performed the Bragg shape phase retrieval. T.A.W. and R.A.K. merged the 3D data. R.F. collected and evaluated the reference data set; R.A.K., T.A.W., J.M.H. and R.F. refined the structure and calculated the electron density maps; and H.N.C., P.F., J.C.H.S. and I.S. wrote the manuscript with discussion and improvements from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.N.C. ([henry.chapman@desy.de](mailto:henry.chapman@desy.de)).

## METHODS

**Experimental set-up.** The experiments were performed at LCLS<sup>4</sup>, at SLAC, at the AMO beamline<sup>29</sup> *in vacuo* using the CAMP end station<sup>12</sup>. X-ray pulses, generated at a repetition rate of 30 Hz, were focused to a spot with a full-width at half-maximum of 7  $\mu\text{m}$  (full-width of 13  $\mu\text{m}$  at 10% maximum irradiance) and a pulse fluence of 900 J cm<sup>-2</sup>, corresponding to a peak power density (irradiance) in excess of 10<sup>16</sup> W cm<sup>-2</sup> at 70-fs duration. The pnCCD detectors were read out, digitized and stored at the 30-Hz rate of the delivered LCLS pulses. Each detector panel consists of 512  $\times$  1,024 pixels 75  $\times$  75  $\mu\text{m}^2$  in area. The rear detectors, located 564 mm from the jet, record low-angle scattering from 0.1° to 4.0° in the vertical scattering plane, and the front detectors, located 68 mm from the jet, cover 4.6° to 40.5° in the same vertical plane. The largest scattering-angle magnitude accepted by the front detector was 47.9°, corresponding to a resolution,  $d$ , of 8.5 Å at a wavelength of 6.9 Å. X-ray fluorescence from the water jet was filtered by an 8- $\mu\text{m}$ -thick polyimide film in front of the pnCCDs.

A liquid microjet<sup>8,9</sup> was used to inject the nanocrystal suspension into the FEL beam at a flow rate of 10  $\mu\text{l min}^{-1}$ . The microjet was emitted from a 40- $\mu\text{m}$ -diameter capillary and focused to a 4- $\mu\text{m}$ -diameter column by a coaxial flow of helium. The X-ray attenuation in the water was at most 30%. The interaction region of the X-rays and crystals is located in the continuous liquid column, before the Rayleigh break-up of the jet into drops, such that most of the X-ray scattering from the liquid is confined to a narrow vertical streak in reciprocal space.

Crystallization conditions of photosystem I nanocrystals were established by determining the phase diagrams<sup>31,32</sup>. Nanocrystals were grown in batches at 10 mg ml<sup>-1</sup> protein concentration (30  $\mu\text{M}$  P700, or 10  $\mu\text{M}$  photosystem I trimer) and low ionic strength (8 mM MgSO<sub>4</sub>, 5 mM MES, pH 6.4, and 0.02%  $\beta$ -dodecylmaltoside) at 4 °C. The photosystem I nanocrystals were then suspended in harvesting buffer (5 mM MES, pH 6.4, and 0.02%  $\beta$ -dodecylmaltoside) to establish a protein concentration of 1 mg ml<sup>-1</sup>. The crystal suspension was filtered through 2- $\mu\text{m}$  cut-off filters (In-line Filter, Upchurch) and stored at 4 °C until use in the experiment.

The nanocrystals are needles of hexagonal cross-section, with the long axis of the needle along the  $c$  axis and an aspect ratio (length to maximum diameter of hexagon) ranging from 1:1 to 2:1, as determined from reconstructing single-shot views of the whole crystal from their shape transforms. For example, Fig. 2a shows a view of the crystal almost perpendicular to the  $c$  axis, where we reconstruct a shape of aspect ratio 1.6:1. A view along the  $c$  axis (Fig. 2c) shows the hexagonal profile. Large, millimetre-sized, crystals of photosystem I have an aspect ratio of up to 5:1, which is seen to decrease with decreasing crystal size.

The nanocrystal suspension was introduced directly into the microjet through a sample loop (Supplementary Fig. 1). A micropore filter in the fluid delivery line was used to restrict the size of the nanocrystals to less than 2  $\mu\text{m}$ . The suspension was diluted to observe a crystal 'hit rate' of 20% (Supplementary Fig. 2), to minimize the occurrence of double hits. The observed concentration of crystals was therefore 0.2 per illuminated volume of 4  $\times$  4  $\times$  13  $\mu\text{m}^3$ , or 10<sup>9</sup> crystals per millilitre. The overall photosystem I protein concentration after dilution was 1 mg ml<sup>-1</sup>, and a complete set of structure factors was obtained from 1,850,000 X-ray pulses, or 10 mg of protein. With the current set-up, at the 30-Hz X-ray pulse rate less than 0.004% of the continuously flowing solution was exposed to the X-ray beam, so only one in 25,000 nanocrystals was actually hit by an X-ray pulse.

Details of the acquisition of diffraction patterns and the primary data reduction are given in Supplementary Methods.

**3D merging of intensities.** Peaks in the processed patterns were located in each pattern using the algorithm of ref. 33, and their locations were mapped into three dimensions according to the curvature of the Ewald sphere, the calibrated detector geometry and the X-ray wavelength. The 3D peak locations for each pattern in turn were presented to the auto-indexing program DirAx<sup>20</sup>. If DirAx succeeded in finding a unit cell for the peaks, linear combinations of the cell basis vectors were checked for correspondence with the photosystem I unit cell<sup>5</sup> from the Protein Data Bank (1D, 1JB0). If a match was found, pixel intensities were summed within a circle of ten-pixel radius centred on the pixel closest to each located Bragg condition. Patterns were rejected if fewer than 10% of the detected peaks were accounted for by unit-cell parameters from DirAx. From 1,850,000 recorded patterns, we identified 112,725 as hits (more than ten detected peaks) and 15,445 were successfully indexed. New peak-finding and -indexing algorithms are under development and are expected to increase significantly the number of patterns that can be indexed, thereby further reducing the number of protein crystals required for a useful data set. The variation of pixel solid angle across the detector plane was accounted for, as was polarization of the X-ray beam assuming complete horizontal polarization. A list of reflection indices and intensities was produced for each individual diffraction pattern, and merging was performed by taking the mean value for the intensity of each unique reflection. Because the indexing algorithm makes use of the positions of the peaks but not their intensities, it was unable to distinguish between crystal orientations related by the symmetry of the lattice. As the symmetry of the lattice is higher than that of the actual structure of photosystem I, an ambiguity exists in that each pattern could correspond to one of two possible orientations. For programmatic convenience, these data (with actual space group symmetry  $P6_3$ ) were merged as  $P6_322$  and treated as though merohedrally twinned during refinement (see below). A 3D rendering of the final full data set is shown in Supplementary Fig. 4.

**Data quality.** Metrics of the merged data quality are shown in Supplementary Table 1 and discussed in Supplementary Information. We carried out a rigid-body refinement of the published photosystem I structure (Protein Data Bank ID, 1JB0) to the merged structure factors using the program REFMAC<sup>34</sup> in twin mode. The refinement  $R$  and  $R_{\text{free}}$  values were 0.25 and 0.23, respectively. The  $2mF_o - DF_c$  electron density map<sup>35</sup> at a resolution of 8.5 Å is shown in Fig. 3c. The corresponding  $2mF_o - DF_c$  electron density map from the conventional synchrotron data, truncated to a resolution of 8.5 Å, is shown in Fig. 3d. The electron density maps show the large subunits PsaA and PsaB, as well as the membrane extrinsic subunits. The transmembrane helices, and even some loop structures, are clearly visible. In these figures, the ribbon representation of the protein model is shown in yellow and the atoms of three iron-sulphur clusters are depicted in red.

- Fromme, P. & Grotjohann, I. in *Membrane Protein Crystallization* (ed. DeLukas, L.) 192–224 (Curr. Top. Membr. 63, Elsevier, 2009).
- Hunter, M. S. *et al.* X-ray diffraction from membrane protein nanocrystals. *Biophys. J.* (in the press).
- Zaefferer, S. New developments of computer-aided crystallographic analysis in transmission electron microscopy. *J. Appl. Cryst.* **33**, 10–25 (2000).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Praznikar, J., Afonine, P. V., Guncar, G., Adams, P. D. & Turk, D. Averaged kick maps: less noise, more signal...and probably less bias. *Acta Crystallogr. D* **65**, 921–931 (2009).



# Single mimivirus particles intercepted and imaged with an X-ray laser

M. Marvin Seibert<sup>1\*</sup>, Tomas Ekeberg<sup>1\*</sup>, Filipe R. N. C. Maia<sup>1\*</sup>, Martin Svenda<sup>1</sup>, Jakob Andreasson<sup>1</sup>, Olof Jönsson<sup>1</sup>, Duško Odić<sup>1</sup>, Bianca Iwan<sup>1</sup>, Andrea Rocker<sup>1</sup>, Daniel Westphal<sup>1</sup>, Max Hantke<sup>1</sup>, Daniel P. DePonte<sup>2</sup>, Anton Barty<sup>2</sup>, Joachim Schulz<sup>2</sup>, Lars Gumprecht<sup>2</sup>, Nicola Coppola<sup>2</sup>, Andrew Aquila<sup>2</sup>, Mengning Liang<sup>2</sup>, Thomas A. White<sup>2</sup>, Andrew Martin<sup>2</sup>, Carl Caleman<sup>1,2</sup>, Stephan Stern<sup>2,3</sup>, Chantal Abergel<sup>4</sup>, Virginie Seltzer<sup>4</sup>, Jean-Michel Claverie<sup>4</sup>, Christoph Bostedt<sup>5</sup>, John D. Bozek<sup>5</sup>, Sébastien Boutet<sup>5</sup>, A. Alan Miahnahri<sup>5</sup>, Marc Messerschmidt<sup>5</sup>, Jacek Krzywinski<sup>5</sup>, Garth Williams<sup>5</sup>, Keith O. Hodgson<sup>6</sup>, Michael J. Bogan<sup>6</sup>, Christina Y. Hampton<sup>6</sup>, Raymond G. Sierra<sup>6</sup>, Dmitri Starodub<sup>6</sup>, Inger Andersson<sup>7</sup>, Saša Bajt<sup>8</sup>, Miriam Barthelmess<sup>8</sup>, John C. H. Spence<sup>9</sup>, Petra Fromme<sup>10</sup>, Uwe Weierstall<sup>10</sup>, Richard Kirian<sup>9</sup>, Mark Hunter<sup>10</sup>, R. Bruce Doak<sup>9</sup>, Stefano Marchesini<sup>11</sup>, Stefan P. Hau-Riege<sup>12</sup>, Matthias Frank<sup>12</sup>, Robert L. Shoeman<sup>13</sup>, Lukas Lomb<sup>13</sup>, Sascha W. Epp<sup>14,15</sup>, Robert Hartmann<sup>16</sup>, Daniel Rolles<sup>13,14</sup>, Artem Rudenko<sup>14,15</sup>, Carlo Schmidt<sup>14,15</sup>, Lutz Foucar<sup>13,14</sup>, Nils Kimmel<sup>17,18</sup>, Peter Holl<sup>16</sup>, Benedikt Rudek<sup>14,15</sup>, Benjamin Erk<sup>14,15</sup>, André Hömke<sup>14,15</sup>, Christian Reich<sup>16</sup>, Daniel Pietschner<sup>17,18</sup>, Georg Weidenspointner<sup>17,18</sup>, Lothar Strüder<sup>14,17,18,19</sup>, Günter Hauser<sup>17,18</sup>, Hubert Gorke<sup>20</sup>, Joachim Ullrich<sup>14,15</sup>, Ilme Schlichting<sup>13,14</sup>, Sven Herrmann<sup>17,18</sup>, Gerhard Schaller<sup>17,18</sup>, Florian Schopper<sup>17,18</sup>, Heike Soltau<sup>16</sup>, Kai-Uwe Kühnel<sup>15</sup>, Robert Andritschke<sup>17,18</sup>, Claus-Dieter Schröter<sup>15</sup>, Faton Krasniqi<sup>13,14</sup>, Mario Bott<sup>13</sup>, Sebastian Schorb<sup>21</sup>, Daniela Rupp<sup>21</sup>, Marcus Adolph<sup>21</sup>, Tais Gorkhover<sup>21</sup>, Helmut Hirsemann<sup>8</sup>, Guillaume Potdevin<sup>8</sup>, Heinz Graafsma<sup>8</sup>, Björn Nilsson<sup>8</sup>, Henry N. Chapman<sup>2,3</sup> & Janos Hajdu<sup>1</sup>

X-ray lasers offer new capabilities in understanding the structure of biological systems, complex materials and matter under extreme conditions<sup>1–4</sup>. Very short and extremely bright, coherent X-ray pulses can be used to outrun key damage processes and obtain a single diffraction pattern from a large macromolecule, a virus or a cell before the sample explodes and turns into plasma<sup>1</sup>. The continuous diffraction pattern of non-crystalline objects permits oversampling and direct phase retrieval<sup>2</sup>. Here we show that high-quality diffraction data can be obtained with a single X-ray pulse from a non-crystalline biological sample, a single mimivirus particle, which was injected into the pulsed beam of a hard-X-ray free-electron laser, the Linac Coherent Light Source<sup>5</sup>. Calculations indicate that the energy deposited into the virus by the pulse heated the particle to over 100,000 K after the pulse had left the sample. The reconstructed exit wavefront (image) yielded 32-nm full-period resolution in a single exposure and showed no measurable damage. The reconstruction indicates inhomogeneous arrangement of dense material inside the virion. We expect that significantly higher resolutions will be achieved in such experiments with shorter and brighter photon pulses focused to a smaller area. The resolution in such experiments can be further extended for samples available in multiple identical copies.

Diffraction studies of crystalline samples have led to spectacular breakthroughs in physics, chemistry and biology over the past hundred years. Many important targets are difficult or impossible to crystallize, and this creates systematic blank areas in the structural sciences. X-ray lasers offer the possibility of stepping beyond X-ray crystallography, to extend structural studies to single, non-crystalline particles or molecules<sup>1</sup>. In this Letter, we present results on biological imaging with

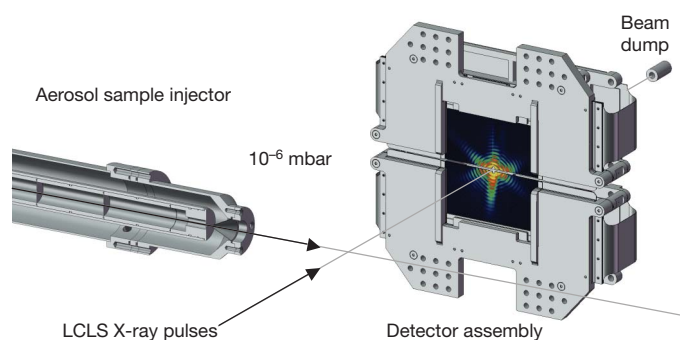
an X-ray free-electron laser, and bring together all the elements required for structural studies of single, non-crystalline objects.

Mimivirus (*Acanthamoeba polyphaga* mimivirus) is the largest known virus<sup>6</sup>. Its size is comparable to the size of the smallest living cells (in fact, the name mimivirus stands for ‘microbe-mimicking virus’). The viral capsid (0.45 µm in diameter) has a pseudo-icosahedral appearance and is covered by an outer layer of dense fibrils<sup>7,8</sup>. The total diameter of the particle, including fibrils, is about 0.75 µm. Mimivirus is too big for a full three-dimensional reconstruction by cryo-electron microscopy<sup>7</sup> and its fibrils prevent crystallization. The genome<sup>9</sup> has 1.2 million base pairs (comparable to a small bacterium) and contains several genes previously thought to be present only in cellular organisms, including components of the protein translation apparatus. Mimivirus can be infected by a smaller virus, named a ‘virophage’<sup>10</sup>, which seems to be the first example of a virus behaving as a parasite of another virus<sup>8</sup>. Studies of mimivirus are causing a paradigm shift in virology and have led to renewed debates about the origin and the definition of viral and cellular life<sup>11</sup>.

Figure 1 shows the experimental arrangement for imaging single virus particles. The sample injector, which uses aerodynamic focusing, was mounted into the CFEL-ASG Multi-Purpose (CAMP) instrument<sup>12</sup> on the Atomic, Molecular and Optical Science (AMO) beam-line<sup>13</sup> at the Linac Coherent Light Source<sup>5</sup> (LCLS). We recorded far-field diffraction patterns at a reduced pressure (10<sup>−6</sup> mbar) to minimize background scattering. Mimivirus was aerosolized from a volatile buffer (250 mM ammonium acetate, pH 7.5) using a gas dynamic nebulizer<sup>14</sup> in a helium atmosphere. The beam of adiabatically cooled virus particles was guided through an aerodynamic lens stack (similar to the one described in ref. 15) and entered the interaction zone with an estimated velocity of 60–100 m s<sup>−1</sup>. The particles were intercepted randomly by

<sup>1</sup>Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden. <sup>2</sup>Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. <sup>3</sup>University of Hamburg, Notkestrasse 85, 22607 Hamburg, Germany. <sup>4</sup>Information Génomique et Structurale, CNRS-UPR2589, Aix-Marseille Université, Institut de Microbiologie de la Méditerranée, Parc Scientifique de Luminy, Case 934, 13288 Marseille Cedex 9, France. <sup>5</sup>LCLS, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>6</sup>PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA. <sup>7</sup>Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala Biomedical Centre, Box 590, S-751 24 Uppsala, Sweden. <sup>8</sup>Photon Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. <sup>9</sup>Department of Physics, PSF470, Arizona State University, Tempe, Arizona 85287-1504, USA. <sup>10</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA. <sup>11</sup>Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>12</sup>Lawrence Livermore National Laboratory, 7000 East Avenue, Mail Stop L-211, Livermore, California 94551, USA. <sup>13</sup>Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. <sup>14</sup>Max Planck Advanced Study Group, Center for Free-Electron Laser Science, Notkestrasse 85, 22607 Hamburg, Germany. <sup>15</sup>Max-Planck-Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany. <sup>16</sup>PN Sensor GmbH, Römerstrasse 28, 80803 München, Germany. <sup>17</sup>Max-Planck-Institut Halbleiterlabor, Otto-Hahn-Ring 6, 81739 München, Germany. <sup>18</sup>Max-Planck-Institut für Extraterrestrische Physik, Giessenbachstrasse, 85741 Garching, Germany. <sup>19</sup>Universität Siegen, Emmy-Noether Campus, Walter Flex Strasse 3, 57068 Siegen, Germany. <sup>20</sup>Forschungszentrum Jülich, Institut ZEL, 52425 Jülich, Germany. <sup>21</sup>Institut für Optik und Atomare Physik, Technische Universität Berlin, Hardenbergstrasse 36, 10623 Berlin, Germany.

\*These authors contributed equally to this work.



**Figure 1 | The experimental arrangement.** Mimivirus particles were injected into the pulse train of the LCLS at the AMO experimental station<sup>13</sup> with a sample injector built in Uppsala. The injector was mounted into the CAMP instrument<sup>12</sup>. The aerodynamic lens stack is visible in the centre of the injector body, on the left. Particles leaving the injector enter the vacuum chamber and are intercepted randomly by the LCLS pulses. The far-field diffraction pattern of each particle hit by an X-ray pulse is recorded on a pair of fast p–n junction charge-coupled device (pnCCD) detectors<sup>12</sup>. The intense, direct beam passes through an opening in the centre of the detector assembly and is absorbed harmlessly behind the sensitive detectors. Some of the low-resolution data also go through this gap and are lost in the current set-up.

the LCLS pulses. The X-ray energy was 1.80 keV (6.9-Å wavelength) and the pulse length was 70 fs (full-duration at half-maximum). The X-ray beam diameter at the interaction point was about 10 µm (full-width at half-maximum), with a maximum of  $1.6 \times 10^{10}$  photons per square micrometre in the centre of this beam. This translates to a peak power density of  $6.5 \times 10^{15}$  W cm<sup>-2</sup>. Forward-scattered diffraction patterns were recorded on a pair of pnCCD detectors<sup>12</sup>. The direct beam exited through an opening between the two detector halves and was absorbed in a beam dump behind the detectors (Fig. 1). The detector pair was placed 564 mm away from the interaction point, giving maximum full-period resolutions of 10.2 nm at the edges and 7.2 nm at the corners of the compound detector at 1.8-keV photon energy.

Figure 2a, b shows single shot X-ray diffraction patterns of individual mimivirus particles, and Fig. 2c shows a transmission electron micrograph of a single mimivirus particle. Each of the diffraction patterns contains about 1,700,000 scattered photons. The lowest-resolution data are missing between the two detector halves, so the total number of scattered photons exceeds this number. Figure 2d, e shows autocorrelation functions calculated from the diffraction patterns. Missing low-resolution data act as a high-pass filter. For an object of extent  $D$ , the extent of its autocorrelation is  $2D$  and the diffraction intensities are band-limited with a Nyquist rate of  $1/2D$ . The size and shape of the autocorrelation functions in Fig. 2d, e are indicative of hits on single virus particles. Figure 2f, g shows the reconstructed exit wavefronts for these mimivirus particles. The shapes and sizes of the reconstructed objects agree with data from prior cryo-electron microscopy studies in which 30,000 images were averaged<sup>7</sup>. In contrast, the reconstructed structures in Fig. 2f, g come from single shots from single particles, and demonstrate the power of this new imaging concept<sup>1</sup>.

We performed image reconstruction by iterative phase retrieval implemented in the Hawk software package<sup>16</sup>, using the RAAR algorithm<sup>17</sup> enhanced with both reality and positivity constraints. The support was handled by a Shrinkwrap algorithm<sup>18</sup> with the constraint of having a specific area that was estimated from the autocorrelation function. Weakly constrained modes in the reconstructions were identified and removed, using the formalism of ref. 19. This is a linear algebra method to compensate for noise, or the lack of constraints in the missing central region of the pattern. The uncertainty in the overall density was less than 10% after the identification and removal of the unconstrained modes. We then fitted these modes to match the total density of a spherical or a suitably rotated icosahedral profile. The missing modes were adjusted to give a total density that best matched the target.

Residual phase fluctuations were then suppressed by averaging many reconstructions, using different random seeds. The results gave improved image reliability. For details, see Methods.

We estimated the image resolution in the reconstruction by computing the phase retrieval transfer function<sup>2,20</sup> (PRTF; Fig. 2h, i), which represents the confidence in the retrieved phases as a function of resolution. No consensus has emerged so far on what single PRTF value should be used as the measure of resolution (values between 0.5 and 0.1 can be found in the literature; see Methods). We characterize resolution by the point where the PRTF drops to  $1/e$  (ref. 20), and this corresponds to a full-period resolution of 32 nm in both cases. We expect significantly higher resolutions in such experiments with shorter and brighter photon pulses focused to a smaller area.

In principle, resolution could reach less than 1 nm in a single exposure with a biological object of similar size to the mimivirus particle<sup>3</sup>. This resolution would require a free-electron laser pulse shorter than about 5 fs at 1.8-keV energy and a photon flux on the sample of more than  $3 \times 10^{11}$  photons per square micrometre<sup>3</sup>. This pulse length and photon flux are beyond the initial capabilities of the LCLS, although there have already been indications of nearly transform-limited LCLS pulses lasting only a few femtoseconds and containing about  $5 \times 10^{11}$  photons per pulse in the unfocused beam<sup>21</sup>.

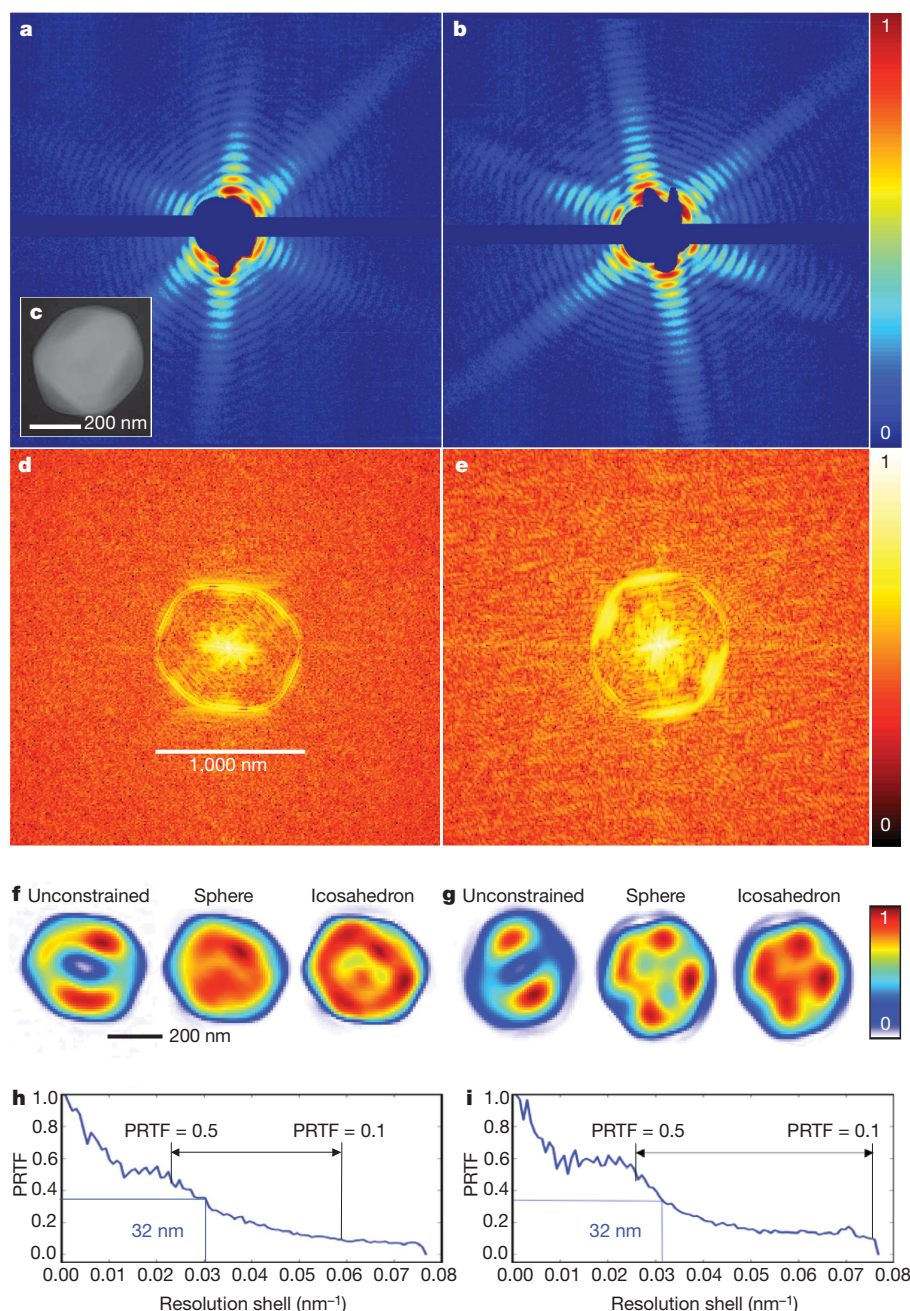
With very short pulses, exposures could be over before there is time for significant Auger emission or for the development of secondary electron cascades in the sample<sup>1</sup>. The conventional handicap of X-rays relative to electrons in imaging could thus be reversed and made into a net gain over a broad range of sample sizes. First experiments at the LCLS show a significant drop in the photoelectric cross-section of hollow atoms<sup>22</sup>. This effect was predicted earlier<sup>1</sup>, but it is larger than expected and can already be measured with LCLS pulses 20–80 fs in duration<sup>22</sup>. The results show photoabsorption decreased 20-fold in hollow neon to equal the cross-section of coherent scattering<sup>22</sup>. In addition, neon ions with double core holes had an extended lifetime<sup>22</sup>. At 1.8-keV photon energy, more than 90% of the total photoelectric cross-section of carbon, nitrogen and oxygen can be attributed to 1s electrons. Ejection of these electrons at the beginning of an intense and short pulse could practically stop photoionization without significantly changing the elastic cross-sections of outer-shell electrons.

We see no measurable sample deterioration. With the X-ray pulses used in this study, the explosion of micrometre-sized objects is hydrodynamic<sup>3</sup> and the sample burns from the outside inwards, rarefying and destroying outer contours first. Trapped electrons move inwards to neutralize an increasingly positive core, and leave behind a positively charged outer layer, which then peels off over some picoseconds<sup>23</sup>. The reconstructed exit wavefront of the mimivirus particle shows well-defined outer contours and gives a sample size consistent with the intact virus capsid (we do not expect to see the thin viral fibrils at the length scales accessible here). Other studies of protein nanocrystals<sup>24</sup> at the LCLS at 0.9-nm resolution show no measurable deterioration of Bragg peaks during illumination with pulses similar to those used here. The size of these protein nanocrystals was similar to the size of the mimivirus particles.

At this stage, it is unclear how reproducible is the interior structure of mimivirus particles (or that of any other viral particles) in terms of atomic positions, and this will need further study. The viral inner capsid consists of a thin protein shell (about 7 nm thick) lined with phospholipid membranes. The structure of the protein shell seems to be reproducible to at least 6.5 nm resolution<sup>7</sup>. Figure 2d, e suggests an inhomogeneous interior structure for the virion. The interior structure does not necessarily follow the pseudo-icosahedral outer shape (the capsid is believed to have a single, five-fold symmetry axis<sup>7</sup>).

The penetration depth of X-rays permits studies on the interiors of large objects. The methods applied here require no modifications to the sample such as staining, freezing, sectioning, radiolabelling or crystallization, and can also be used to image cells that are alive at the time of the exposure. The amount of missing data can be reduced by adding an additional detector pair behind the first pair. Another





**Figure 2 | Single-shot diffraction patterns on single virus particles give interpretable results.** **a, b**, Experimentally recorded far-field diffraction patterns (in false-colour representation) from individual virus particles captured in two different orientations. **c**, Transmission electron micrograph of an unstained Mimivirus particle, showing pseudo-icosahedral appearance<sup>7</sup>. **d, e**, Autocorrelation functions for **a** (**d**) and **b** (**e**). The shape and size of each autocorrelation correspond to those of a single virus particle after high-pass filtering due to missing low-resolution data. **f, g**, Reconstructed images after iterative phase retrieval with the Hawk software package<sup>16</sup>. The size of a pixel corresponds to 9 nm in the images. Three different reconstructions are shown for each virus particle: an averaged reconstruction with unconstrained Fourier

modes<sup>19</sup> and two averaged images after fitting unconstrained low-resolution modes to a spherical or an icosahedral profile, respectively. The orientation of the icosahedron was determined from the diffraction data. The results show small differences between the spherical and icosahedral fits. **h, i**, The PRTF for reconstructions where the unconstrained low-resolution modes were fitted to an icosahedron. All reconstructions gave similar resolutions. We characterize resolution by the point where the PRTF drops to 1/e (ref. 20). This corresponds to 32-nm full-period resolution in both exposures. Arrows mark the resolution range with other cut-off criteria found in the literature (Methods). Resolution can be substantially extended for samples available in multiple identical copies<sup>1,25–28</sup>.

necessary improvement is to increase the dynamic range of the detectors. In our experiments, there were shots extending to significantly higher resolutions than those reported here but they contained too many saturated pixels at low angles (more missing modes), preventing image reconstruction. With reproducible samples, where the experiment can be repeated on a new object, a three-dimensional data set can be collected, and the resolution extended (even from weak individual

exposures) by merging redundant data<sup>25–29</sup>. Studies of virus particles with higher-intensity photon pulses and improved detectors could answer the question of whether the core is reproducible to subnanometre resolution or whether the viral genome has the ‘molecular individualism’ that genomic DNA structures explore *in vitro*<sup>30</sup>. *Note added in proof:* In a previous study<sup>31</sup>, synchrotron radiation was used to obtain X-ray diffraction data on a herpes virus.



## METHODS SUMMARY

Experiments were performed with the CAMP instrument<sup>12</sup> on the AMO beamline<sup>13</sup> at the LCLS<sup>5</sup>, with the LCLS running at a repetition rate of 30 Hz. CAMP supports a variety of imaging and atomic/molecular physics experiments.

Diffraction patterns were recorded on a pair of pnCCD detectors<sup>12</sup> (maximum read-out speed, 250 frames per second). The sample-to-detector distance was 564 mm. The active area of each detector half was 76.8 mm × 38.4 mm and contained 1,024 × 512 pixels of area 75 × 75 μm<sup>2</sup>. The full-well capacity of a pixel was 280,000 electrons, corresponding to ~570 X-ray photons per pixel at 1.8-keV photon energy.

The electron bunch was 70 fs long (full-duration at half-maximum), but the corresponding photon bunch is thought to be shorter<sup>22</sup>. The photon bunch contained  $8 \times 10^{11}$  photons per pulse (0.24 mJ at 1.8 keV) and had a diameter of ~10 μm (full-width at half-maximum) at the interaction point, giving ~1.6 × 10<sup>10</sup> photons per square micrometre in the centre of the beam and a peak power density of ~6.5 × 10<sup>15</sup> W cm<sup>-2</sup>. Background scattering from residual gas in the vacuum chamber did not exceed the read-out noise of the detectors nor the noise of the diffuse photon background (<1.3 photons per pixel). This is remarkable, considering that the number of photons in the pulse was nearly 100,000,000,000 times higher than the photon background.

Purified mimivirus was transferred into a volatile buffer and the suspension was aerosolized with helium gas in a gas dynamic nebulizer<sup>14</sup>. The aerosol of hydrated virus particles was sampled into a differentially pumped injector through an inlet nozzle coupled to a skimmer. The aerosol (in helium atmosphere) passed through a variable relaxation chamber from where the equilibrated and adiabatically cooled particles entered a differentially pumped aerodynamic lens stack. Particles focused by the aerodynamic lens were intercepted randomly by the LCLS pulses. Diffraction patterns of free-flying virus particles were exceptionally clean.

Image reconstruction was performed with the open source Hawk software<sup>16</sup>, available from <http://xray.bmc.uu.se/hawk>. The background-corrected diffraction patterns and the Hawk configuration files used in the reconstructions can be downloaded from this site.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 16 July; accepted 9 December 2010.**

- Neutze, R. *et al.* Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
- Chapman, H. N. *et al.* Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Phys.* **2**, 839–843 (2006).
- Bergh, M. *et al.* Feasibility of imaging living cells at sub-nanometer resolution by ultrafast X-ray diffraction. *Q. Rev. Biophys.* **41**, 181–204 (2008).
- Nagler, B. *et al.* Turning solid aluminium transparent by intense soft X-ray photoionization. *Nature Phys.* **5**, 693–696 (2009).
- Emma, P. *et al.* First lasing and operation of an ångström-wavelength free-electron laser. *Nature Photon.* **4**, 641–647 (2010).
- La Scola, B. *et al.* A giant virus in amoebae. *Science* **299**, 2033 (2003).
- Xiao, C. *et al.* Structural studies of the giant mimivirus. *PLoS Biol.* **7**, e1000092 (2009).
- Claverie, J. M. & Abergel, C. Mimivirus and its virophage. *Annu. Rev. Genet.* **43**, 49–66 (2009).
- Raoult, D. *et al.* The 1.2-megabase genome sequence of mimivirus. *Science* **306**, 1344–1350 (2004).
- La Scola, B. *et al.* The virophage as a unique parasite of the giant mimivirus. *Nature* **455**, 100–104 (2008).
- Claverie, J. M. & Abergel, C. Mimivirus: the emerging paradox of quasi-autonomous viruses. *Trends Genet.* **26**, 431–437 (2010).
- Strüder, L. *et al.* Large-format, high-speed, X-ray pnCCDs combined with electron and ion imaging spectrometers in a multipurpose chamber for experiments at 4th generation light sources. *Nucl. Instrum. Methods Phys. Res. A* **614**, 483–496 (2010).
- Bozek, J. D. AMO instrumentation for the LCLS X-ray FEL. *Eur. Phys. J. Spec. Top.* **169**, 129–132 (2009).
- DePonte, D. P. *et al.* Gas dynamic virtual nozzle for generation of microscopic droplet streams. *J. Phys. D* **41**, 195505 (2008).
- Bogan, M. J. *et al.* Single particle X-ray diffractive imaging. *Nano Lett.* **8**, 310–316 (2008).
- Maia, F. R. N. C., Ekeberg, T., van der Spoel, D. & Hajdu, J. Hawk: the image reconstruction package for coherent X-ray diffractive imaging. *J. Appl. Crystallogr.* **43**, 1535–1539 (2010).

- Luke, D. R. Relaxed averaged alternating reflections for diffraction imaging. *Inverse Probl.* **21**, 37–50 (2005).
- Marchesini, S. *et al.* X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B* **68**, 140101 (2003).
- Thibault, P., Elser, V., Jacobsen, C., Shapiro, D. & Sayre, D. Reconstruction of a yeast cell from X-ray diffraction data. *Acta Crystallogr. A* **62**, 248–261 (2006).
- Shapiro, D. *et al.* Biological imaging by soft X-ray diffraction microscopy. *Proc. Natl Acad. Sci. USA* **102**, 15343–15346 (2005).
- Ding, Y. *et al.* Measurements and simulations of ultralow emittance and ultrashort electron beams in the Linac Coherent Light Source. *Phys. Rev. Lett.* **102**, 254801 (2009).
- Young, L. *et al.* Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature* **466**, 56–61 (2010).
- Hau-Riege, S. P. *et al.* A sacrificial tamper slows down sample explosion in flash diffraction experiments. *Phys. Rev. Lett.* **104**, 064801 (2010).
- Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* doi:10.1038/nature09750 (this issue).
- Huldt, G., Szoke, A. & Hajdu, J. Diffraction imaging of single particles and biomolecules. *J. Struct. Biol.* **144**, 219–227 (2003).
- Fung, R. *et al.* Structure from fleeting illumination of faint spinning objects in flight. *Nature Phys.* **5**, 64–67 (2009).
- Loh, N. T. D. & Elser, V. Reconstruction algorithm for single-particle diffraction imaging. *Phys. Rev. E* **80**, 026705 (2009).
- Bortel, G., Faigel, G. & Tegze, M. Classification and averaging of random orientation single macromolecular diffraction patterns at atomic resolution. *J. Struct. Biol.* **166**, 226–233 (2009).
- Maia, F. R. N. C., Ekeberg, T., Timneanu, N., van der Spoel, D. & Hajdu, J. Structural variability and the incoherent addition of scattered intensities in single-particle diffraction. *Phys. Rev. E* **80**, 031905 (2009).
- Perkins, T. T., Smith, D. E. & Chu, S. Single polymer dynamics in an elongated flow. *Science* **276**, 2016–2021 (1997).
- Song, C. *et al.* Quantitative imaging of single, unstained viruses with coherent X rays. *Phys. Rev. Lett.* **101**, 158101 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the following agencies: the Swedish Research Councils; the Swedish Foundation for International Cooperation in Research and Higher Education; Stiftelsen Olle Engkvist Byggmästare; the Swedish University of Agricultural Sciences; the Helmholtz Association (VH-VI-302); the DFG Cluster of Excellence at the Munich Centre for Advanced Photonics; the Centre National de la Recherche Scientifique; Agence Nationale de la Recherche (ANR-BLAN08-0089); the Hamburg Ministry of Science and Research and Joachim Herz Stiftung, as part of the Hamburg Initiative for Excellence in Research (LEXI); the Hamburg School for Structure and Dynamics; the Max Planck Society, the US National Science Foundation (grants MCB 0919195 and MCB-1021557); and the US Department of Energy, through the PULSE Institute. Portions of this research were carried out at the Linac Coherent Light Source, a National User Facility operated by Stanford University on behalf of the US Department of Energy, Office of Basic Energy Sciences. We are grateful to B. Hedman and N. Timneanu for their help and to the scientific and technical staff of the LCLS for their outstanding facility and support.

**Author Contributions** J.H. and H.N.C. conceived the experiment. C.A., V.S., J.-M.C., M.S., O.J., A. Rocker, I.A. and D.O. prepared and characterized the samples. J.D.B., C.B. and K.O.H. created the LCLS beamline. J.H., H.N.C., J.S., L.G., A.B., N.C., A.A., A.M., J.K., K.O.H. and S.P.H.-R. developed the imaging concept. S.W.E., R.H., D. Rolles, A. Rudenko, C.S., L.F., N.K., P.H., B.R., B.E., A.H., Ch.R., D.P., G.W., L.S., G.H., H. Gorke, J.U., I.S., S.H., G.S., F.S., H.S., K.-U.K., R.A., C.-D.S., F.K., M. Bott, S. Schorb, D. Rupp, M.A., T.G., H.H., L.G., G.P., H. Graafsma and B.N. designed and set up the CAMP instrument and/or developed and operated the pnCCD detectors. S. Bajt and M. Barthelmeß coordinated instrumentation, worked on engineering and prepared filters and calibration samples. J.K., S.P.H.-R., A.B., H.N.C., J.S., A.M. and N.C. characterized the focus. B.I., D.W., M.S., M.M.S. and J.H. built the sample injector. D.P.D., J.H., M.S., D.W., U.W., R.K., M. Hunter, R.B.D. and J.C.H.S. designed and built the nanospray nebulizer. J.H., M.S., M.M.S., F.R.N.C.M., J.A., A.A.M., A. Rocker, M.J.B., C.Y.H., R.G.S., S. Boutet, I.A., O.J., D.S., A.B., J.S., D.P.D., A.A., M.L., J.K., T.A.W., A.M., R.L.S., L.L., M. Barthelmeß, J.C.H.S., P.F., I.S., U.W., R.K., M. Hunter, R.B.D., M.F., G.W., M. Bott, S.W.E., B.E., L.F., R.H., N.K., L.L., D. Rupp, B.R., A. Rudenko, R.L.S., L.S., I.S., C.S., J.U. and H.N.C. characterized the imaging apparatus and carried out the experiment. T.E., F.R.N.C.M., M. Hantke, M.M.S., A.B., T.A.W., A.M., A.A., S.M., M.M. and J.H. analysed the data. T.E., F.R.N.C.M., A.B. and S.M. performed image reconstructions. J.H. and T.E. wrote the manuscript with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.H. ([janos.hajdu@xray.bmc.uu.se](mailto:janos.hajdu@xray.bmc.uu.se)) or C.A. ([chantal.abergel@igs.cnrs-mrs.fr](mailto:chantal.abergel@igs.cnrs-mrs.fr)).

## METHODS

**Experimental set-up at the LCLS.** Experiments were performed with the CAMP instrument<sup>12</sup> on the AMO beamline<sup>13</sup> at the LCLS<sup>5</sup>, with the LCLS running at a repetition rate of 30 Hz. CAMP supports a variety of imaging and atomic/molecular physics experiments.

Diffraction patterns were recorded on a pair of pnCCD detectors<sup>12</sup> (maximum read-out speed, 250 frames per second). The sample-to-detector distance was 564 mm. The active area of each detector half was 76.8 mm × 38.4 mm and contained 1,024 × 512 pixels of area 75 × 75 μm<sup>2</sup>. The full-well capacity of a pixel was 280,000 electrons, corresponding to ~570 X-ray photons per pixel at 1.8-keV photon energy.

The electron bunch was 70 fs long (full-duration at half-maximum), but the corresponding photon bunch is thought to be shorter<sup>22</sup>. The photon bunch contained  $8 \times 10^{11}$  photons per pulse (0.24 mJ at 1.8 keV) and had a diameter of ~10 μm (full-width at half-maximum) at the interaction point, giving  $\sim 1.6 \times 10^{10}$  photons per square micrometre in the centre of the beam and a peak power density of  $\sim 6.5 \times 10^{15}$  W cm<sup>-2</sup>. Background scattering from residual gas in the vacuum chamber did not exceed the read-out noise of the detectors nor the noise of the diffuse photon background (<1.3 photons per pixel). This is remarkable, considering that the number of photons in the pulse was nearly 100,000,000,000 times higher than the photon background.

Purified mimivirus<sup>32</sup> particles were transferred into a volatile buffer (250 mM ammonium acetate, pH 7.5) and the suspension ( $10^{12}$  particles per millilitre) was aerosolized at a rate of about 5 μl min<sup>-1</sup>, using helium gas in a gas dynamic nebulizer<sup>14</sup>. The aerosol of hydrated virus particles was sampled into a differentially pumped sample injector through an inlet nozzle coupled to a skimmer. Most of the nebulizing gas, and vapours of the volatile buffer, were pumped away at this point. The heavier aerosol (in a wet helium atmosphere at a pressure of about 10<sup>-2</sup> mbar) passed through a variable-volume relaxation chamber from where the equilibrated and adiabatically cooled particles entered a differentially pumped aerodynamic lens stack. The pressure dropped from about 10<sup>-2</sup> mbar to about 10<sup>-4</sup> mbar at the exit of the lens. Particles focused by the aerodynamic lens entered the interaction zone (10<sup>-6</sup> mbar) with an estimated velocity of 60–100 m s<sup>-1</sup> and were intercepted randomly by the LCLS pulses.

Data processing included removal of signal from known bad or saturated pixels, correction for the residual common mode offsets and application of a flat-field correction. The corrected patterns were used directly without symmetrization.

Image reconstruction was performed with the open-source Hawk software package<sup>16</sup>, using the RAAR algorithm<sup>17</sup> and its support constraint, enhanced by additional reality and positivity constraints. Hawk is available from <http://xray.bmc.uu.se/hawk>. The background-corrected diffraction patterns and the Hawk configuration files used in the reconstructions can be downloaded from this site.

A Fourier constraint was applied to match Fourier amplitudes with experimental amplitudes through a projection. No explicit Fourier constraints were used for regions of missing data, although these regions were implicitly constrained in Fourier space by the real-space constraints. The support was handled using a Shrinkwrap algorithm<sup>18</sup> with the constraint of having a specific area that was estimated from the autocorrelation function.

Weakly constrained modes in the reconstructions were identified and removed, using the formalism of ref. 19. This is a linear algebra method to compensate for noise, or the lack of constraints in the missing central region of the pattern. The diffracted amplitudes in the region of missing data can be recovered by iterative phasing algorithms, but for patterns where this region is extensive the recovered amplitudes will be unreliable<sup>19</sup>. Missing modes were identified and their constraining power was calculated by performing a singular-value decomposition on the transform from the region of missing data to the support. The singular values identify the modes that are most weakly constrained, as the singular vectors, and determine their constraining power. In the patterns discussed in this paper, there are modes with very low constraining power. These modes are therefore virtually unconstrained and their strength had to be estimated in another way. The threshold used for identifying unconstrained modes was 0.999, corresponding to a constraining power of 0.045. The uncertainty in the total image density dropped to less than 10% after removing these modes. Missing modes were fitted to match the total density of a spherical or a suitably rotated icosahedral profile. The missing modes were adjusted to give a total density that best matched the target.

The number of weakly constrained modes that were classified as unconstrained differs slightly between the two reconstructions because the support recovered through the Shrinkwrap algorithm<sup>17</sup> is slightly different for Fig. 2a and Fig. 2b. For the reconstruction starting from Fig. 2a, the median number of missing modes was 8 and the average was 7.75 modes. For the reconstruction starting from Fig. 2b, the number was slightly higher: median 12 and average 11.85. This difference is due to a larger area being missing in the centre of Fig. 2b owing to there being more saturated pixels.

Residual phase fluctuations were suppressed by averaging many reconstructions, using different random seeds. The results gave improved image reliability. For reconstructions from Fig. 2a, 10,000 iterations were used and 200 reconstructions were obtained from different random starting positions and then used to calculate the PRTF. The support was updated every 20 iterations. All 200 reconstructions had a Fourier error below a threshold of 0.33. For reconstructions from Fig. 2b, 40,000 iterations were performed and 94 reconstructions obtained. Of these, only 56 reconstructions had a Fourier error below a threshold of 0.33. The differences underline the deleterious effect of missing low-resolution data on image reconstruction. Reliable image reconstruction needs a more efficient way of measuring low-angle diffraction data, including a wider dynamic range for the detector. An attenuator disk centred on the X-ray beam and placed over the middle part of the pnCCD detector pair could reduce the strong forward-scattered signal in the middle of the diffraction pattern and bring low-resolution data within the useful dynamic range of the detector.

We estimate the image resolution in the reconstruction by computing the PRTF<sup>22,20</sup>. No consensus has emerged so far on what single PRTF value should be used as the measure of resolution. Values between 0.5 and 0.1 can be found in the literature<sup>33</sup>. We characterize resolution by the point where the PRTF drops to 1/e (ref. 20). Diffraction data extend to higher resolution than the resolution given by the PRTF.

The angle spanned by the signal was small enough for the entire particle to fit within the depth of field. Defocus effects are therefore avoided by using a reality constraint. A resolution of 1 nm at the same X-ray wavelength would require measuring at high angle, leading to significant deviation from the projection image. Real-value constraints would not work in the latter case, and this would make the reconstruction more challenging but by no means impossible (see, for example, ref. 34).

Transmission electron microscopy was performed with a Hitachi H-7100 electron microscope on unstained mimivirus particles deposited on Formvar-coated gold grids.

**A route for improvements.** There is a clear need to achieve higher resolution in single shots. This requires an increased photon flux on the sample as well as a wider dynamic range for detecting photons in the diffraction pattern. The LCLS<sup>5</sup> is capable of delivering very short X-ray pulses<sup>21</sup> to outrun significant sample explosion with more photons per pulse. Tighter focusing and an increased photon output from the LCLS have already been achieved and will increase the flux on the samples in forthcoming runs. A broad dynamic range in detecting photons is necessary to avoid saturation at low angles. In this first set of experiments, there were already exposures with significantly higher resolutions than those reported here, but these exposures contained too many saturated pixels at low angles, preventing image reconstruction. Reliable image reconstruction needs a more efficient way of measuring low-angle diffraction data. A graded attenuator around the central hole of the pnCCD detector pair could help here. An additional pair of detectors placed far behind the first detector pair could record more of the low-angle data over a larger area. Maintaining sample integrity during injection is a key requirement in the experiment. More data on a diverse set of samples (such as cells, viruses and macromolecules) will be needed to map out the available parameter space. Hit rates could be increased by improved injection methods, using a narrower particle beam. A future extension to imaging single macromolecules will need these developments.

32. Byrne, D. *et al.* The polyadenylation site of Mimivirus transcripts obeys a stringent 'hairpin rule'. *Genome Res.* **19**, 1233–1242 (2009).
33. Steinbrener, J. *et al.* Data preparation and evaluation techniques for X-ray diffraction microscopy. *Opt. Express* **18**, 18598–18614 (2010).
34. Seibert, M. M. *et al.* Femtosecond diffractive imaging of biological cells. *J. Phys. At. Mol. Opt. Phys.* **43**, 194015 (2010).

# Late Holocene methane rise caused by orbitally controlled increase in tropical sources

Joy S. Singarayer<sup>1</sup>, Paul J. Valdes<sup>1</sup>, Pierre Friedlingstein<sup>1,2</sup>, Sarah Nelson<sup>1</sup> & David J. Beerling<sup>3</sup>

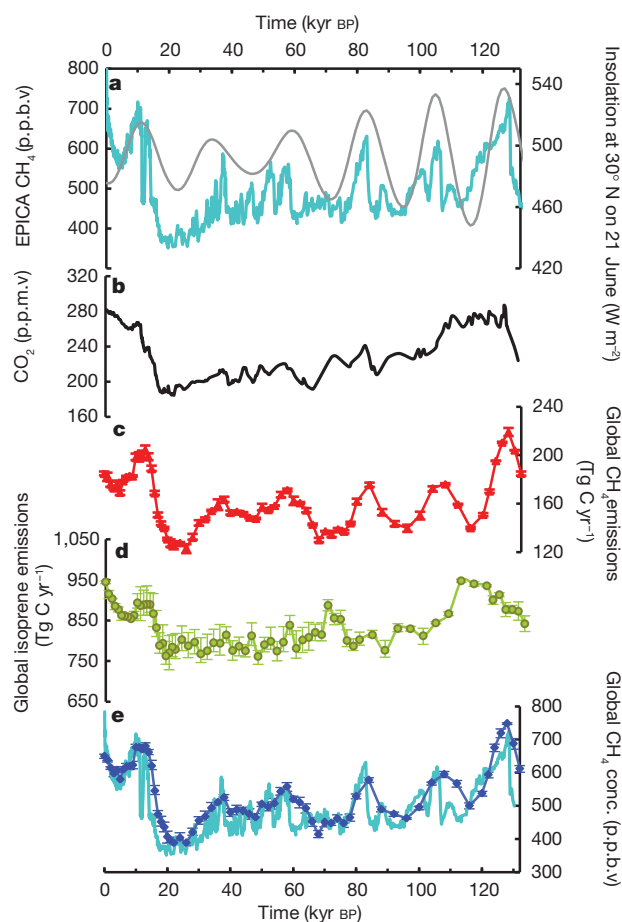
Considerable debate surrounds the source of the apparently 'anomalous'<sup>1</sup> increase of atmospheric methane concentrations since the mid-Holocene (5,000 years ago) compared to previous interglacial periods as recorded in polar ice core records<sup>2</sup>. Proposed mechanisms for the rise in methane concentrations relate either to methane emissions from anthropogenic early rice cultivation<sup>1,3</sup> or an increase in natural wetland emissions from tropical<sup>4</sup> or boreal sources<sup>5,6</sup>. Here we show that our climate and wetland simulations of the global methane cycle over the last glacial cycle (the past 130,000 years) recreate the ice core record and capture the late Holocene increase in methane concentrations. Our analyses indicate that the late Holocene increase results from natural changes in the Earth's orbital configuration, with enhanced emissions in the Southern Hemisphere tropics linked to precession-induced modification of seasonal precipitation. Critically, our simulations capture the declining trend in methane concentrations at the end of the last interglacial period (115,000–130,000 years ago) that was used to diagnose the Holocene methane rise as unique. The difference between the two time periods results from differences in the size and rate of regional insolation changes and the lack of glacial inception in the Holocene. Our findings also suggest that no early agricultural sources are required to account for the increase in methane concentrations in the 5,000 years before the industrial era.

Atmospheric methane (CH<sub>4</sub>) is a strong greenhouse gas influenced by various natural (for example, wetlands, biomass burning) and anthropogenic (for example, rice agriculture, enhanced biomass burning, landfill) sources<sup>7</sup>. The major sink of methane is oxidation in the troposphere by reaction with hydroxyl radicals. Atmospheric hydroxyl radicals are consumed by oxidation of volatile organic compounds (VOCs), including isoprene. Consequently, biogenic VOC fluxes indirectly influence the lifetime of methane in the atmosphere (currently  $8.7 \pm 1.3$  years; ref. 7) by altering the concentration of hydroxyl radicals.

Since the start of the industrial era, atmospheric methane concentrations have more than doubled to 1,780 parts per billion by volume (p.p.b.v.; ref. 7), whereas over the last four glacial cycles (420,000 years; 420 kyr), polar ice core records indicate a natural range of ~700 p.p.b.v. (interglacials) to ~360 p.p.b.v. (glacials)<sup>2</sup> (Fig. 1a). Multi-millennial variations correlate strongly with orbital precession, owing to the predominant response of tropical wetland emissions to low-latitude Northern Hemisphere summer insolation via monsoons<sup>8</sup>. However, a particularly intriguing feature of these ice-core records is that the correlation between methane concentrations and Northern Hemisphere summer insolation (30° N, 21 June) appears to break down from the mid-Holocene onwards, when Northern Hemisphere summer insolation decreases as the atmospheric methane concentration slowly increases<sup>9</sup> from ~550 p.p.b.v. at 5 kyr before present (BP) to ~675 p.p.b.v. shortly before the industrial era (Fig. 1a and 2).

Several mechanisms have been proposed for this 'anomalous' late Holocene trend. First, that emissions from circum-Arctic wetlands increased in a slow response to warmer, stable Holocene conditions<sup>5,10</sup>,

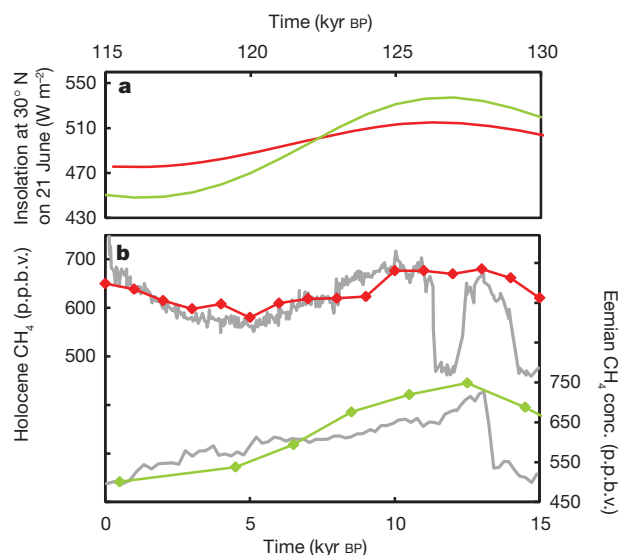
a mechanism supported by  $\delta^{13}\text{CH}_4$  measurements from Greenland ice cores<sup>6,11</sup>. Second, changes in the inter-polar gradient of methane<sup>4,12</sup> suggested increased natural tropical emissions. However, Northern Hemisphere tropical wetland regions are believed to have experienced a natural drying since the early Holocene<sup>13,14</sup>. Third, a controversial hypothesis posits that it results from early agricultural activity, particularly human-formed wetlands created by irrigating rice paddies, farming ruminants, burning biomass and human waste<sup>1,3</sup>. It remains



**Figure 1 | Time series of model and ice core data for the last glacial cycle.** **a**, Atmospheric CH<sub>4</sub> concentrations from the Antarctic EPICA Dome C ice core<sup>2</sup> (aqua) and 30° N 21 June insolation<sup>26</sup> (grey). **b**, Composite atmospheric CO<sub>2</sub> concentrations from Antarctic ice cores. **c**, Modelled global annual total methane emissions from experiment ALL. **d**, Global annual isoprene emissions from ALL. **e**, Modelled methane concentrations from ALL (blue diamonds) compared with the EPICA ice core record (aqua). Error bars in **c–e** denote the range of values obtained when the vegetation/wetland model was run with three different mean climatologies taken from the HadCM3 ALL simulations.

<sup>1</sup>Bristol Research Initiative for the Dynamic Global Environment, School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, UK. <sup>2</sup>College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK. <sup>3</sup>Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.





**Figure 2 | A comparison of model- and data-based  $\text{CH}_4$  concentrations from the current interglacial and the previous interglacial (Eemian) periods.** **a**, Northern Hemisphere summer insolation at  $30^\circ \text{N}$  for the Holocene (red line) and Eemian (green line). **b**, EPICA ice core  $\text{CH}_4$  record (grey line) and modelled  $\text{CH}_4$  concentrations for the Holocene (red line and symbols), and similarly for the Eemian (green line and symbols).

uncertain whether agricultural activities and population were sufficient to influence global atmospheric composition by 5 kyr before present (BP)<sup>3</sup>. However, although the idea of significant early anthropogenic impacts on atmospheric methane levels has been questioned on several fronts<sup>15</sup>, no modelling studies have yet been able to explain both the Holocene and pre-Holocene methane record.

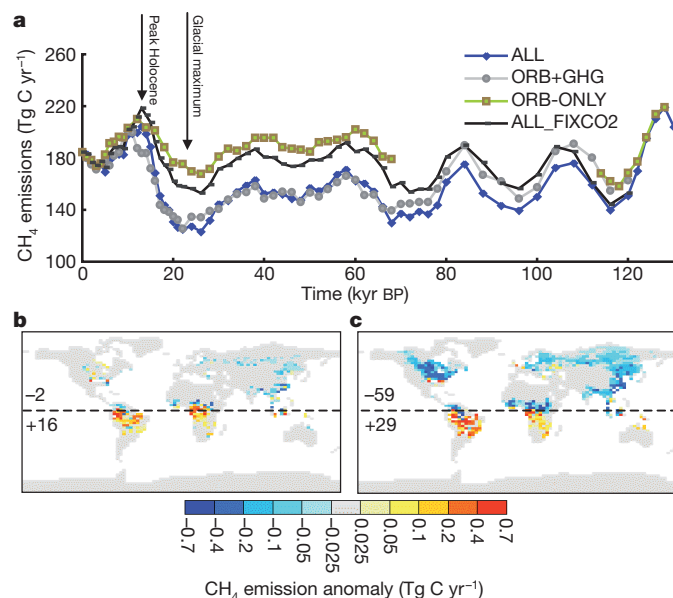
Here we report model-based reconstructions of terrestrial wetland methane emissions and atmospheric concentrations over the last glacial cycle (130 kyr) due only to natural forcing mechanisms. Specifically, our study addresses the issues of (1) whether natural processes explain the late-Holocene methane increase and the pre-Holocene methane record, (2) the main location of increased methane emissions during the late Holocene, and (3) the relative importance of different natural processes in producing glacial–interglacial methane changes. We simulated 65 ‘snapshots’ spanning 130 kyr of the last glacial cycle<sup>16</sup> using the coupled ocean–atmosphere Hadley Centre climate model (HadCM3). The resulting climatologies were used to drive the Sheffield Dynamic Global Vegetation Model (SDGVM)<sup>17</sup> coupled to a wetlands methane emission model<sup>18</sup> to predict the location of vegetation, wetlands, methane emissions and VOC emissions over this period. Modelled emissions were used to reconstruct atmospheric methane concentrations for the last glacial cycle using a simplified scheme that captures the relevant atmospheric chemistry, drawing on earlier chemistry-model experiments<sup>18</sup>. We undertook three sets of experiments designed to identify mechanisms of atmospheric methane increase: (1) varying orbital configuration only (ORB-ONLY), (2) varying orbital configuration and atmospheric greenhouse gas ( $\text{CO}_2$ ,  $\text{CH}_4$  and  $\text{N}_2\text{O}$ ) concentrations (ORB+GHG), and (3) varying orbital configuration, greenhouse gas concentrations, ice-sheet extent and sea level (ALL). See Methods for further details.

Variations in modelled global methane emissions over the last glacial cycle exhibit several robust features which allow us to exclude specific candidate mechanisms for the late Holocene rise in methane. Methane emissions from experiment ALL display a high degree of correlation with  $30^\circ \text{N}$  summer insolation over the full glacial cycle ( $r^2 = 0.77$ , excluding the late Holocene), reminiscent of the ice core methane record (Fig. 1a and c). These results also reproduce an early Holocene decrease in methane emissions followed by a smaller increase during

the late Holocene. Given that the model contains no prescribed human emissions, this increase suggests that the increases in methane concentrations seen in the ice core record are predominantly natural.

Modelled methane concentrations (Fig. 1e) demonstrate considerable variation at the precessional scale and display a strong increase from 5 kyr BP onwards in good agreement with the ice core record. This is primarily due to increases in wetland emissions, with a much smaller contribution ( $\sim 20\%$ ) from isoprene variations (Supplementary Fig. 4). Crucially, the model simulates both the decrease in methane concentrations following the Eemian interglacial maximum and the late Holocene increase (Fig. 2). The difference between the Eemian and Holocene methane trends has previously been used as key evidence in support of the idea that early agriculture had a significant impact on atmospheric methane concentrations in the past 5 kyr (ref. 15). Here, we simulate both the Eemian and Holocene methane concentration variations without anthropogenic influence.

Several processes could potentially be driving the methane rise during the late Holocene. We used sensitivity simulations to separate the effects of orbital forcing, changes in atmospheric  $\text{CO}_2$  concentration, ice-sheet distribution, and sea level. Between 8 kyr BP and shortly before the industrial era ( $\sim 0.2$  kyr BP), atmospheric  $\text{CO}_2$  increased by  $\sim 17$  p.p.m.v. (Fig. 1b). This increase could influence wetland methane emissions in two ways. First, it could induce warmer, wetter conditions, which would promote increased plant primary productivity and hence soil carbon available for methanogenesis. Second, it could result in higher methane emissions by increasing primary production through  $\text{CO}_2$  fertilization<sup>17</sup>. In a separate experiment, we forced SDGVM with the climate output of experiment ALL but with a prescribed pre-industrial  $\text{CO}_2$  concentration (280 p.p.m.v.). This methodology isolates the effect of radiative  $\text{CO}_2$  impacts by eliminating  $\text{CO}_2$  fertilization. The results (ALL\_FIXCO2) demonstrate that  $\text{CO}_2$  fertilization contributes  $\sim 1/3$  of the range of Last Glacial Maximum to peak Holocene variations in experiment ALL methane emissions (Fig. 3a). However, even after removing  $\text{CO}_2$  fertilization effects, the rise in late Holocene methane emissions remains.



**Figure 3 | Temporal and spatial patterns of modelled methane emission changes for the last glacial cycle.** **a**, Time series of modelled global annual total  $\text{CH}_4$  emissions from the four experiments: ALL, ALL\_FIXCO2, ORB+GHG and ORB-ONLY. **b**, Anomalies in annual methane emissions (pre-industrial minus 5 kyr BP) for simulation ALL. **c**, Anomalies in annual methane emissions for the Eemian interglacial (116 kyr BP) for simulation ALL. Dashed line in **b** and **c** represents the Equator, and numbers above and below are the anomalies in methane emissions for the northern and southern hemisphere in  $\text{Tg C yr}^{-1}$ .

The ORB-ONLY experiment eliminates both radiative and fertilization impacts of CO<sub>2</sub>. It exhibits the smallest variation in global methane emissions of all the experiments (Fig. 3a), with glacial–interglacial difference approximately 50% of that in experiment ALL. As in the other experiments, there is a strong correlation with Northern Hemisphere summer insolation. Importantly, during the Holocene, ORB-ONLY methane emission changes are of similar magnitude and sign as in the other experiments: the late Holocene increase is evident, even though all other climate driving forces have been fixed. Therefore, whether the slight increase in CO<sub>2</sub> from the early to late Holocene is natural<sup>19</sup> or anthropogenic<sup>1</sup>, it is not the main driving mechanism of late Holocene methane. The results of the sensitivity experiments imply strongly that forcing of climate by changes in insolation played a significant role in the late Holocene methane increase. Slight variation occurs in the timing of the late Holocene increase (between 5 kyr and 3 kyr BP) among the different experiments (Fig. 3a), which arises partly from variations in the boundary conditions between experiments, and also from inherent variability in the climatologies used to drive the SDGVM. The uncertainty from climate variability was quantified by running the SDGVM with three different 30-year average climatologies from each simulation to produce repeat values of methane and isoprene emissions. The full range of resultant values was used as the uncertainty estimate (Fig. 1c–e). The rise in methane during the late Holocene remains significant after these uncertainties are taken into consideration.

We next address the mechanisms by which the last 5 kyr appear to deviate from the correlation of methane emissions with orbital precessional variation<sup>1</sup>. Examination of the simulated spatial distribution of anomalies (value at 0 kyr BP minus value at mid-Holocene; Fig. 3b) shows that the Southern Hemisphere is the main location of increasing wetland emissions during the late Holocene. During this period there is negligible change in ice-sheet volume or atmospheric CO<sub>2</sub> due to relatively small variations in orbital configuration compared to the previous interglacial. In the ORB-ONLY experiment, methane emissions lead Northern Hemisphere summer insolation in general by 3–5 kyr (Supplementary Fig. 5b), owing to the balance between the large methane sources (Eurasia and South America), which have different seasonal cycles of precipitation and wetland emissions (Eurasia restricted to June to August, whereas South America emissions extend over November to May; see Supplementary Fig. 7). However, when glacial–interglacial variation is simulated through additional prescription of ice-sheets and CO<sub>2</sub> changes, these influence the relative magnitudes of emissions from the different regions and their seasonality; the changes over time in the proportion of emissions from different regions change the phase of global methane variation subtly, depending on the degree of glaciation and size of insolation variation. In the last glacial, the methane emissions from South America are smaller and much less variable on the precessional timescale. By comparison, emissions from Eurasia and East Asia (Supplementary Fig. 2) decrease, but maintain a strong variation with insolation. For further information, see Supplementary Information. The balance of these regional factors results in the overall correlation of global methane emissions with summer insolation at 30° when all forcings are included.

The reason why this correlation appears to break down in the mid-Holocene but not in the Eemian is due to the lack of glacial inception exerting subtle changes in the strength of the different source regions. The regional anomalies (value at 0 kyr minus value at 5 kyr BP) in experiment ALL show that the Southern Hemisphere increases in methane emissions are larger than the concurrent decreases in the Northern Hemisphere (Fig. 3b). Anomalies for time slices of comparable orbital configuration in the Eemian (value at 116 kyr minus value at 122 kyr BP) show that the increases in the Southern Hemisphere are larger than those in the Holocene in line with the larger changes in insolation, but the increases in the Southern Hemisphere are outweighed by the decline from Northern Hemisphere sources (Fig. 3c) due to the difference in latitudinal variation in insolation. The decreases at high northern latitudes are amplified in experiment ALL compared

to ORB-ONLY (which leads 30° N summer insolation in the Eemian as well as the Holocene) due to glacial inception at 116 kyr BP. The complexity of the responses to the various regional forcings means that concentrations cannot be derived from one insolation curve, and that an ‘earth system’ modelling approach employed here is essential to understand drivers of Pleistocene methane records.

The climate and vegetation models used in this study have been extensively validated against modern observations, as well as palaeoproxy records in the case of HadCM3 (Online Methods). We have also quantified statistical uncertainties related to internal variability of the climate model for experiment ALL and demonstrated that the findings are robust. However, within the scope of this study we have not been able to quantify uncertainties arising from the physical parameterization values in the models. We recognize several other caveats associated with our approach. For example, there is no millennial-scale variation, although this is not essential for the longer timescale we are considering. The modelling approach neglects transient responses, such as the potential of boreal wetlands to respond to delayed circum-Arctic warming in the Holocene<sup>6</sup>. Given the agreement between simulated and measured methane concentrations obtained throughout the glacial cycle, we propose that including such effects is not crucial. The isoprene model is realistic but empirical in terms of its dependence on carbon dioxide<sup>20</sup>, and there is uncertainty in the mechanisms involved<sup>21</sup>.

We conclude that the late Holocene increase in methane can be primarily ascribed to increasing emissions from the Southern Hemisphere tropics. In the Holocene, unlike the last interglacial, these increases are not counteracted by equivalent decreases in Northern Hemisphere emissions. We suggest therefore that direct anthropogenic influences are not necessary to explain the late Holocene methane record.

## METHODS SUMMARY

We have performed multiple snapshot simulations with the Hadley Centre coupled atmosphere–ocean climate model, HadCM3<sup>22,23</sup>. Three sets of snapshot simulations were conducted, consisting of 65 model runs covering the past 130 kyr, at a frequency ranging from every 4,000 years at the start (120–80 kyr BP), to every 2,000 years (80–22 kyr BP) and to every 1,000 years at the end (22 kyr BP to the ‘pre-industrial’ 0 kyr BP)<sup>16</sup>. Experiment ALL has glacial–interglacial changes to ice-sheet volume derived from the ICE5G model<sup>24</sup>, greenhouse gases<sup>2,25</sup> and orbital configuration<sup>26</sup>. Experiment ORB+GHG includes changes to orbital configuration and greenhouse gases, and ORB-ONLY has only changes to orbital configuration (other forcings fixed at pre-industrial magnitudes). In all simulations the initial conditions were based on a spun-up pre-industrial simulation and each was run for 500 years. The results presented here are mean climatologies of the last 30 years of each simulation.

SDGVM was used to simulate patterns of net primary production (NPP), leaf area index (LAI) and plant functional types using the HadCM3-derived monthly inputs of temperature, precipitation, relative humidity and cloudiness, and global data sets of soil texture<sup>17,27</sup>. Seasonal patterns of wetland methane emissions from terrestrial ecosystems were computed with a process-based model describing the dependence of anaerobic microbial methane production and aerobic oxidation on temperature, gross primary productivity, soil respiration and soil water table depth<sup>18,28</sup>. Monthly trace gas VOC emissions (isoprene, monoterpene and others) are calculated using a global scheme<sup>29</sup> coupled to SDGVM modified to account for monthly LAI changes. CO<sub>2</sub> dependence of isoprene emissions was parameterized using an empirically based function<sup>20</sup>. Atmospheric methane concentrations were calculated by a first order approximation using the experiments from a previous modelling study<sup>18</sup> which estimated the sensitivity of global methane concentrations to isoprene and methane emissions.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 8 June; accepted 2 December 2010.

1. Ruddiman, W. F. The anthropogenic greenhouse era began thousands of years ago. *Clim. Change* **61**, 261–293 (2003).
2. Spahni, R. *et al.* Atmospheric methane and nitrous oxide of the Late Pleistocene from Antarctic ice cores. *Science* **310**, 1317–1321 (2005).
3. Ruddiman, W. F., Guo, Z., Zhou, X., Wu, H. & Yu, Y. Early rice farming and anomalous methane trends. *Quat. Sci. Rev.* **27**, 1291–1295 (2008).

4. Chappellaz, J. *et al.* Variations of the Greenland/Antarctic concentration difference in atmospheric methane during the last 11,000 years. *J. Geophys. Res.* **102**, 15987–15997 (1997).
5. Schmidt, G. A., Shindell, D. T. & Harder, S. L. A note on the relationship between ice core methane concentrations and insolation. *Geophys. Res. Lett.* **31**, L23206, doi:10.1029/2004GL021083 (2004).
6. Sowers, T. Atmospheric methane isotope records covering the Holocene period. *Quat. Sci. Rev.* **29**, 213–221 (2010).
7. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 130–234 (Cambridge Univ. Press, 2007).
8. Loulergue, L. *et al.* Orbital and millennial-scale features of atmospheric CH<sub>4</sub> over the past 800,000 years. *Nature* **453**, 383–386 (2008).
9. Ruddiman, W. F. & Thomson, J. S. The case for human causes of increased atmospheric CH<sub>4</sub> over the last 5000 years. *Quat. Sci. Rev.* **20**, 1769–1777 (2001).
10. Blunier, T. *et al.* Variations in atmospheric methane concentration during the Holocene epoch. *Nature* **374**, 46–49 (1995).
11. Fischer, H. *et al.* Changing boreal methane sources and constant biomass burning during the last termination. *Nature* **452**, 864–867 (2008).
12. Brook, E. J. *et al.* On the origin and timing of rapid changes in atmospheric methane during the last glacial period. *Glob. Biogeochem. Cycles* **14**, 559–572 (2000).
13. Kutzbach, J. E. Monsoon climate of the Early Holocene: climate experiment with the Earth's orbital parameters for 9000 years ago. *Science* **214**, 59–61 (1981).
14. Cohmap members. Climatic changes of the last 18,000 years: observations and model simulations. *Science* **241**, 1043–1052 (1988).
15. Ruddiman, W. F. The early anthropogenic hypothesis: challenges and responses. *Rev. Geophys.* **45**, RG4001 (2007).
16. Singarayer, J. S. & Valdes, P. J. High-latitude climate sensitivity to ice-sheet forcing over the last 120 kyr. *Quat. Sci. Rev.* **29**, 43–55 (2010).
17. Beerling, D. J. & Woodward, F. I. *Vegetation and the Terrestrial Carbon Cycle: Modelling the First 400 Million Years* (Cambridge Univ. Press, 2001).
18. Valdes, P. J., Beerling, D. J. & Johnson, C. E. The ice age methane budget. *Geophys. Res. Lett.* **32**, L02704, doi:10.1029/2004GL021004 (2005).
19. Stocker, B., Strassmann, K. & Joos, F. Sensitivity of Holocene atmospheric CO<sub>2</sub> and the modern carbon budget to early human land use: analyses with a process-based model. *Biogeosci. Discuss.* **7**, 921–952 (2010).
20. Possell, M., Hewitt, C. N. & Beerling, D. J. The effects of glacial atmospheric CO<sub>2</sub> concentrations and climate on isoprene emissions by vascular plants. *Glob. Change Biol.* **11**, 60–69 (2005).
21. Arneeth, A. *et al.* Why are estimates of global terrestrial isoprene emissions so similar (and why is this not so for monoterpenes)? *Atmos. Chem. Phys.* **8**, 4605–4620 (2008).
22. Gordon, C. *et al.* The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16**, 147–168 (2000).
23. Pope, V. D., Gallani, M. L., Rowntree, P. R. & Stratton, R. A. The impact of new physical parameterisations in the Hadley Centre climate model: HadAM3. *Clim. Dyn.* **16**, 123–146 (2000).
24. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annu. Rev. Earth Planet. Sci.* **32**, 111–149 (2004).
25. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
26. Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
27. Woodward, F. I., Smith, T. M. & Emanuel, W. R. A global primary productivity and phytogeography model. *Glob. Biogeochem. Cycles* **9**, 471–490 (1995).
28. Cao, M. K. S., Marshall, S. & Gregson, K. Global carbon exchange and methane emissions from natural wetlands: application of a process-based model. *J. Geophys. Res.* **101**, 14399–14414 (1996).
29. Guenther, A. *et al.* A global model of natural volatile organic compound emissions. *J. Geophys. Res.* **100**, 8873–8892 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** J.S.S. and P.J.V. thank the BBC for commissioning the climate simulations for their TV series, *The Incredible Human Journey*. Thanks are also due to I. Woodward and M. Lomas for earlier SDGVM development. P.J.V. and D.J.B. acknowledge additional support through Royal Society-Wolfson Research Merit Awards and the Leverhulme Trust. J.S.S. thanks R. Bailey for comments.

**Author Contributions** J.S.S. and P.J.V. performed the climate model simulations. P.J.V. performed initial analysis of methane emissions from experiment ALL. J.S.S. performed the main analysis, sensitivity experiments and led the writing of the paper with contributions from all other authors. P.F. and D.J.B. contributed expertise concerning methane emissions and the carbon cycle. D.J.B. and S.N. evolved the development of the vegetation and methane models.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.S.S. ([joy.singarayer@bristol.ac.uk](mailto:joy.singarayer@bristol.ac.uk)).



## METHODS

**HadCM3 experiments.** We have performed multiple 'snapshot' simulations with the Hadley Centre climate model, HadCM3. HadCM3 consists of coupled dynamic ocean<sup>22</sup>, atmosphere<sup>23</sup> and sea-ice models. The sea-ice model includes parameterizations of ice drift and leads<sup>30</sup>. The resolution of the atmospheric model is  $2.5^\circ$  in latitude by  $3.75^\circ$  in longitude by 19 unequally spaced levels in the vertical. The spatial resolution over the ocean in HadCM3 is  $1.25^\circ$  by  $1.25^\circ$  by 20 unequally spaced layers in the ocean extending to a depth of 5,200 m. In this version of the model, interactive vegetation is not included. The HadCM3 model has been validated against modern observations<sup>22,23</sup>. The model has been compared to other climate models within the Paleoclimate Model Intercomparison Project (PMIP)<sup>31</sup> and PMIP2<sup>32,33</sup> under pre-industrial, mid-Holocene (6 kyr BP) and Last Glacial Maximum (21 kyr BP) conditions. In a previous study we have also compared the glacial cycle snapshot simulations used here against ice core palaeo-records<sup>16</sup>. The global glacial–interglacial temperature range compares well with other models and the palaeo-record<sup>16,32</sup>. Temperature changes directly over Greenland and East Antarctica are underestimated compared to ice core reconstructions<sup>16</sup>, a bias that is common to all global climate models. Low-latitude variation in precipitation in the inter-tropical convergence zone is similar to other models<sup>31,33</sup>, with a small northerly bias in West Africa<sup>33</sup>.

We have performed three sets of snapshot simulations, each consisting of 65 model runs covering the whole of the past 130 kyr (ref. 16), at a frequency ranging from every 4,000 years at the start of the period (between 120 kyr BP and 80 kyr BP), to every 2,000 years from 80 kyr BP to 22 kyr BP and to every 1,000 years from 22 kyr BP to the 0 kyr BP. The 0 kyr time slice has 'pre-industrial' boundary conditions for ~1850 AD. Experiment ALL has glacial–interglacial changes to ice-sheet volume and extent, greenhouse gases and orbital configuration<sup>26</sup>. Experiment ORB+GHG includes changes to orbital configuration and greenhouse gases, and ORB-ONLY has only changes to orbital configuration, with all other forcings fixed at pre-industrial magnitudes (for this experiment time slices were run for 0–70 kyr BP and 114–130 kyr BP). Atmospheric concentrations of CO<sub>2</sub> were taken from EDC96<sup>34</sup> (0–22 kyr BP), Taylor Dome<sup>35</sup> (22–62 kyr BP) and Vostok<sup>25</sup> (62–363 kyr BP) ice cores, and CH<sub>4</sub> and N<sub>2</sub>O were taken from EPICA<sup>2</sup> ice cores; all are on the EDC3 timescale<sup>36</sup>. We have developed our ice sheet reconstructions using the ICE5G model<sup>24</sup>. This data set includes a detailed evolution of the ice thickness, extent, and continental isostatic rebound for the whole period from the LGM to the modern at 500-year intervals. We used an anomaly-based method to calculate our palaeogeographic boundary conditions. Anomalies of a particular time-slice palaeogeography minus pre-industrial ICE-5G data are then added to our model pre-industrial geographical boundary conditions<sup>16</sup>.

In all simulations the initial conditions were the same, based on a spun-up pre-industrial simulation. Each simulation was run for 500 years. The approach of using pre-industrial initial conditions for all simulations is a reasonable compromise necessary to enable us to run the simulations simultaneously, meaning it took months rather than years to complete the total set. The results presented here are climatologies of the last thirty years of each simulation.

**SDGVM vegetation and wetland model.** SDGVM simulates global patterns of net primary production (NPP), leaf area index (LAI) and the distribution of plant functional types from monthly inputs of temperature, precipitation, relative humidity and cloudiness, and global data sets of soil texture<sup>17,27</sup>. Core modules of net photosynthesis, stomatal conductance, canopy transpiration, uptake of mineralized nitrogen and responses of these attributes to changes in soil water supply are detailed, and rigorously evaluated against field observations<sup>17,27,37</sup>. A key feature of the model is the coupling of above- and below-ground carbon and nitrogen cycles. Litter production influences soil C and N pools via the Century soil nutrient cycling model<sup>38</sup>, which in turn feedback to influence above-ground primary production.

Local and global-scale predictions of NPP, LAI and plant functional type distribution by the SDGVM have been extensively and successfully evaluated against a wide range of measurements, field observations and satellite products<sup>24,37,39</sup>. SDGVM simulated geographical distribution of plant functional types is in close agreement with maps of 'actual' vegetation<sup>17,27</sup>. Global terrestrial net primary production for the contemporary climate and CO<sub>2</sub> is estimated with SDGVM to be 62 Gt C yr<sup>-1</sup>, in agreement with satellite-based estimate<sup>37</sup> of ~55–60 Gt C yr<sup>-1</sup>. Sensitivity of NPP predictions of SDGVM to CO<sub>2</sub> and climate are similar to those of other dynamic vegetation models<sup>40,41</sup>. The CO<sub>2</sub> fertilization response of NPP compares favourably to that reported in Free Air Carbon Dioxide Enrichment (FACE) experiments for temperate forested sites<sup>42</sup>.

Seasonal patterns of wetland CH<sub>4</sub> emissions from terrestrial ecosystems were computed with a process-based model describing the dependence of anaerobic microbial CH<sub>4</sub> production and aerobic oxidation on temperature, vegetation activity (gross primary productivity), soil respiration and soil water table depth<sup>18,28</sup>. The CH<sub>4</sub> model is coupled to SDGVM and HadCM3 on a monthly time-step. Wetland CH<sub>4</sub> emissions were modified to include the effects of orography by scaling with a linear function of sub-grid orographic variance<sup>18</sup>. The coupled wetland model produces regional and latitudinally averaged CH<sub>4</sub> fluxes similar

to those of ref. 28, and have been validated against observations. Biomass burning from a fire module within SDGVM contributes to total CH<sub>4</sub> emissions. This assumes 80% of the above-ground carbon and nitrogen are lost by fire when the litter content reaches critical dryness<sup>18</sup>. CH<sub>4</sub> emissions from termites and oceans are assumed to be the same of those in the pre-industrial but with altered distributions according to the land-sea areas. These procedures probably mean that the contribution of biogenic CH<sub>4</sub> fluxes from these sources are conservative.

The wetland model produces monthly trace gas VOC emissions (isoprene, monoterpene and other VOCs) using a global scheme<sup>29</sup>, modified to account for monthly LAI changes<sup>18</sup>. Soil biogenic NO<sub>x</sub> fluxes are calculated with an empirical model<sup>43</sup> describing their dependence on temperature, precipitation, and canopy deposition. Lightning NO<sub>x</sub> emissions were calculated based on the convective precipitation amount. To account for recent, although uncertain, evidence of a CO<sub>2</sub> dependence of the emission rate of isoprene<sup>44</sup>, we have used a function to describe this which was estimated from empirical measurements in figure 6 of ref. 20, to modify global isoprene emissions.

**Methane concentration calculations.** In the interests of feasible computation expense, atmospheric methane concentrations were calculated by a first order approximation which used the sensitivity experiments from a previous modelling study<sup>18</sup> to estimate the sensitivity of global methane concentrations to isoprene/monoterpene (which influence the concentrations of the methane sink, hydroxyl radicals) and methane emissions. This study used the atmospheric-only version of the Hadley Centre model (HADAM3<sup>23</sup>) to model climate for pre-industrial and LGM. SDGVM was used to reconstruct vegetation, wetland methane emissions, biomass burning and VOCs, as well as other related variables, and the atmospheric chemistry model STOCHEM<sup>45</sup> was used to reconstruct atmospheric methane concentrations. Subsequently, the chemistry model was run with LGM conditions except for pre-industrial methane emissions or isoprene/monoterpene emissions, or pre-industrial climate to assess the sensitivity of atmospheric methane concentrations and lifetime to each factor in turn. Here we use these sensitivities and assume linear relationships between methane concentrations and emissions of methane and isoprene. Atmospheric methane concentrations are calculated as in equation (1):

$$\text{CH}_4(t) \approx \text{CH}_4(\text{PI}) + A(\text{CH}_4\text{-E}(\text{PI}) - \text{CH}_4\text{-E}(t)) + B(\text{ISO}(\text{PI}) - \text{ISO}(t)) \quad (1)$$

Here CH<sub>4</sub>(*t*) is global methane concentration at time *t* (kyr BP); CH<sub>4</sub>-E(*t*) is global methane emission rate; ISO(*t*) is isoprene emission rate; sensitivity of methane concentrations to methane emissions is  $A = 3.17 \text{ p.p.b.v. per Tg C yr}^{-1}$ , sensitivity of methane concentrations to isoprene emissions is  $B = 0.26 \text{ p.p.b.v. per Tg yr}^{-1}$ .

30. Cattle, H. & Crossley, J. Modelling Arctic climate-change. *Phil. Trans. Royal Soc. Lond. A* **352**, 201–213 (1995).
31. Braconnot, P., Joussaume, S., de Noblet, N. & Ramstein, G. Mid-Holocene and Last Glacial Maximum African monsoon changes as simulated within the Paleoclimate Modelling Intercomparison Project. *Glob. Planet. Change* **26**, 51–66 (2000).
32. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. *Clim. Past* **3**, 261–277 (2007).
33. Braconnot, P. *et al.* Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 2: feedbacks with emphasis on the location of the ITCZ and mid- and high latitudes heat budget. *Clim. Past* **3**, 279–296 (2007).
34. Monnin, E. *et al.* Atmospheric CO<sub>2</sub> concentrations over the last glacial termination. *Science* **291**, 112–114 (2001).
35. Indermühle, A. *et al.* Atmospheric CO<sub>2</sub> concentration from 60 to 20 kyr BP from the Taylor Dome ice core, Antarctica. *Geophys. Res. Lett.* **27**, 735–738 (2000).
36. Parrenin, F. *et al.* The EDC3 age scale for the EPICA Dome C ice core. *Clim. Past* **3**, 485–497 (2007).
37. Beerling, D. J., Woodward, F. I., Lomas, M. & Jenkins, A. J. Testing the responses of a dynamic global vegetation model to environmental change: a comparison of observations and predictions. *Glob. Ecol. Biogeogr. Lett.* **6**, 439–450 (1997).
38. Parton, W. J. *et al.* Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Glob. Biogeochem. Cycles* **7**, 785–809 (1993).
39. Running, S. W. *et al.* A continuous satellite-derived measure of global terrestrial primary production. *Bioscience* **54**, 547–560 (2004).
40. Sitch, S. *et al.* Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Glob. Change Biol.* **14**, 2015–2039 (2008).
41. Hickler, T. *et al.* CO<sub>2</sub> fertilization in temperate FACE experiments not representative of boreal and tropical forests. *Glob. Change Biol.* **14**, 1531–1542 (2008).
42. Woodward, F. I. & Kelly, C. K. Responses of global plant diversity capacity to changes in carbon dioxide concentration and climate. *Ecol. Lett.* **11**, 1229–1237 (2008).
43. Yienger, J. & Levy, H. Empirical model of global soil and biogenic NO<sub>x</sub> emissions. *J. Geophys. Res.* **100**, 11447–11464 (1995).
44. Arneth, A. *et al.* CO<sub>2</sub> inhibition of global terrestrial isoprene emissions: potential implications for atmospheric chemistry. *Geophys. Res. Lett.* **34**, L18813 (2007).
45. Collins, W. J., Stevenson, D., Johnson, C. & Derwent, R. Role of convection in determining the budget of odd hydrogen in the upper troposphere. *J. Geophys. Res.* **104**, 26927–26941 (1999).

# Alternative stable states explain unpredictable biological control of *Salvinia molesta* in Kakadu

Shon S. Schooler<sup>1</sup>, Buck Salau<sup>2</sup>, Mic H. Julien<sup>1</sup> & Anthony R. Ives<sup>3</sup>

Suppression of the invasive plant *Salvinia molesta* by the salvinia weevil is an iconic example of successful biological control. However, in the billabongs (oxbow lakes) of Kakadu National Park, Australia, control is fitful and incomplete. By fitting a process-based nonlinear model to thirteen-year data sets from four billabongs, here we show that incomplete control can be explained by alternative stable states<sup>1–4</sup>—one state in which salvinia is suppressed and the other in which salvinia escapes weevil control. The shifts between states are associated with annual flooding events. In some years, high water flow reduces weevil populations, allowing the shift from a controlled to an uncontrolled state; in other years, benign conditions for weevils promote the return shift to the controlled state. In most described ecological examples, transitions between alternative stable states are relatively rare, facilitated by slow-moving environmental changes, such as accumulated nutrient loading or climate change<sup>5,6</sup>. The billabongs of Kakadu give a different manifestation of alternative stable states that generate complex and seemingly unpredictable dynamics. Because shifts between alternative stable states are stochastic, they present a potential management strategy to maximize effective biological control: when the domain of attraction to the state of salvinia control is approached, augmentation of the weevil population or reduction of the salvinia biomass may allow the lower state to trap the system.

Awareness and concern about the ecological consequences of alternative stable states is growing as more examples have been identified<sup>1,5,7–9</sup>. In many examples, the alternative states are very stable, so that in the absence of an extraordinary perturbation, the system remains at its 'natural' state. Concern arises because states can change abruptly even when the environmental drivers responsible for the change occur gradually; if the ecological system is perturbed by a slow-moving driver, the system may remain largely unchanged until it reaches a threshold catastrophe and abruptly shifts to another state<sup>1</sup>. At present, there is a theoretical enterprise to identify the early warning signs of these abrupt shifts<sup>10,11</sup>. Equally disturbing, once the shift occurs the system will show hysteresis<sup>1</sup>; even if the environmental perturbation were reversed, the system would stay at its new state, inhibiting the ability of managers to repair the system to its desired state<sup>12</sup>.

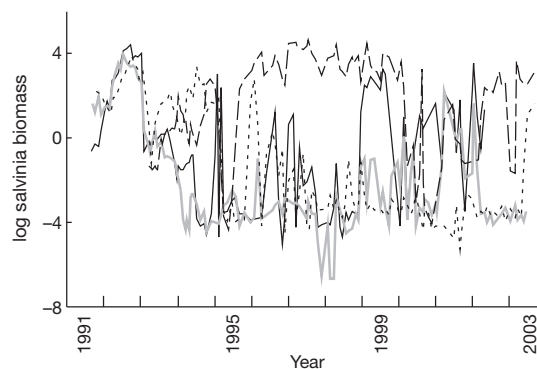
Ecological systems, however, are subjected to stochastic and cyclic perturbations, and if alternative states are weakly stable and perturbations are large enough, then shifts between states may be routine<sup>1,13</sup>. Alternative stable states may thus generate underlying forces that govern the stochastic dynamics of the system, leading to complex and seemingly irregular, eruptive behaviour. In fact, it may be difficult to identify the alternative stable states, yet at the same time be difficult to understand the dynamics of the system without first identifying that alternative stable states exist.

*Salvinia molesta*, a South American aquatic plant, is one of the most widespread and environmentally, economically and socially destructive invasive plant species. Since 1939, it has invaded lake and river systems in tropical and subtropical habitats around the world<sup>14</sup>. Its success is owing to its ability to double in biomass every 3–4 days, and to regenerate

vegetatively even after severe damage or drying<sup>15,16</sup>. It is capable of forming dense mats up to 1-m thick that make waterways unnavigable and displace aquatic organisms<sup>14</sup>. Nonetheless, highly successful biological control is often provided by the salvinia weevil (*Cyrtobagous salviniae*, Curculionidae) that since 1980 has been introduced into most regions where salvinia has invaded<sup>17</sup>. The salvinia weevil is a strict specialist on salvinia; adults feed on growing meristematic tissue (buds), whereas larvae tunnel through vascular tissues<sup>14</sup>, which together often lead to marked reductions in salvinia with no additional expenditure of resources<sup>14,17</sup>.

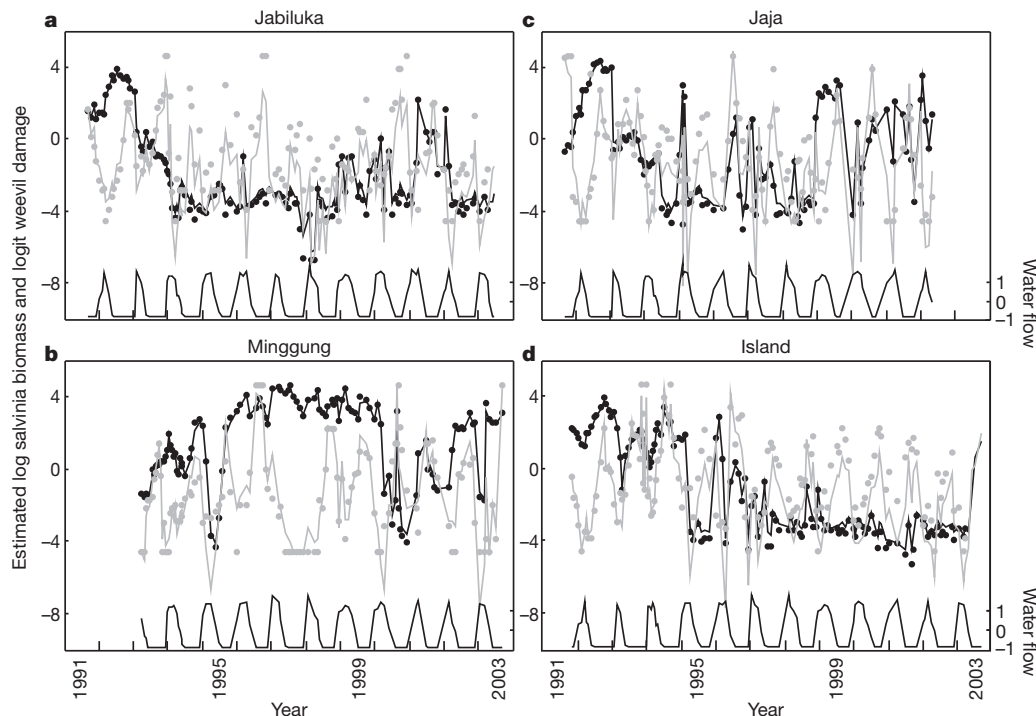
Salvinia invaded Kakadu National Park in 1983<sup>16</sup>, and the salvinia weevil was released later that year<sup>18</sup>. Although the weevils rapidly established and successfully controlled salvinia for several years, in 1988–1990 salvinia resurged to form thick mats. This led to an intensive research project conducted by the Australian Commonwealth Scientific and Industrial Research Organization (CSIRO) in Kakadu<sup>18</sup> and the establishment of a long-term monthly sampling program (Fig. 1). Unlike lake systems that experience continuously successful biological control of salvinia, the Kakadu billabongs are subjected to annual flooding that flushes salvinia downstream, mixing it among billabongs within the same floodplain. Floods also translocate salvinia to and from the billabongs and moist terrestrial sites; salvinia persists in these terrestrial sites during the dry season, sometimes at high biomass, where it has a refuge from the strictly aquatic weevils. Salvinia also occurs in the understory of grasses that grow over water along billabong edges where it is partially protected from both flooding and weevils<sup>14</sup>. Thus, whereas the billabongs of Kakadu are highly perturbed, salvinia has refuges against both flooding and weevils from which it can invade open water.

Fieldwork during 1991–1994 led to the following hypothesis for the failure of continuous biological control. When salvinia is at low density, it has relatively high nitrogen content and high growth rate. Salvinia is thus highly susceptible to biological control, and the many developing



**Figure 1 | Log biomass of salvinia in four Kakadu billabongs.** Grey line shows data for Jabiluka, dashed line for Minggung, black line for Jaja and dotted line for Island. The vertical axis is scaled to have mean zero across all billabongs.

<sup>1</sup>CSIRO Ecosystem Sciences, Long Pocket Laboratories, Indooroopilly, Queensland 4068, Australia. <sup>2</sup>Department of the Environment, Water, Heritage, and the Arts, Kakadu National Park, Jabiru, Northern Territory 0886, Australia. <sup>3</sup>Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706, USA.



**Figure 2 | Fitted model to log salvinia biomass (black) and logit weevil damage (grey) for four billabongs. a–d.** Dots give the raw data, and lines give the updated values from the Kalman filter (Box 1). Data and fitted model are standardized to have mean zero across all billabongs. The log water flow

measured at monitoring stations in each drainage is given by the line at the bottom of the figure and is standardized to have mean zero and variance one (note different axis).

buds support high weevil numbers<sup>19</sup>. Conversely, when salvinia is at high density, growth rates are relatively low, and new buds are scarce. In this condition much of the plant biomass occurs as vegetative, non-meristematic tissue often piled high above the water surface. This salvinia is much less suitable for weevils and is thus difficult to control. Although not stated as such, this is a hypothesis of alternate stable states determined by the growth state of salvinia: a low biomass state maintained by weevil attack, and a high biomass state that has escaped weevil control. As found in other cases of alternative stable states, the salvinia–weevil system involves a species that has distinct life forms<sup>1,20</sup> and a herbivore that can potentially lose control of its food plant<sup>21,22</sup>.

Standard analyses of the time-series data from the four billabongs reveal none of the dynamical hallmarks of successful biological control: there is little evidence that weevil damage is associated with declines in salvinia biomass (Supplementary Information). Further, changes in salvinia biomass between monthly samples are characterized by frequent small steps and occasional large jumps, indicating that nonlinear processes are driving salvinia dynamics (Supplementary Information). Therefore, we built a nonlinear model from what we know about the biology of the system (Box 1). Briefly, the model assumes that salvinia biomass is divided into two categories, one with copious buds that is vulnerable to weevils and the other that is not, with the proportion represented by these two categories depending on total salvinia biomass<sup>18</sup>. The growth of salvinia biomass depends on the amount of vulnerable tissue. Weevils attack this vulnerable tissue non-uniformly, so that attacks can be aggregated among buds. The population growth rate of the weevils depends on the amount of salvinia biomass damaged. There is net migration of salvinia into billabongs from external areas or the grass understory along billabong edges, and mortality/flushing of both salvinia and weevils that depends on the water flow through the drainages. There is also stochastic variation that affects the per capita growth rates of both salvinia and weevils, and this variation can increase with increasing water flow. The latter property accounts for possible increases in unpredictable flushing or filling of billabongs with salvinia or weevils during flood events. The model fits the data well, with most

parameters showing statistically strong effects on the observed dynamics (Fig. 2 and Supplementary Information).

Key biological insights from the fitted parameter values of the model include the following (Table 1). The proportion of salvinia biomass in the category that is invulnerable to weevil attack,  $g$ , is generally high, ranging from  $g_{\min} = 0.91$  to  $g_{\max} = 0.94$ , which is consistent with weevils attacking meristematic tissues (adults) and vascular tissue (larvae). Flooding events reduce the abundance of weevils ( $d_w < 0$ ), yet have little net effect on the mean abundance of salvinia ( $d_s = 0$ ). However, flooding events increase the variability in sample-to-sample fluctuations in salvinia biomass ( $\sigma_{s2} > 0$ ; see Table 1). These two patterns could be caused by flooding events moving salvinia among billabongs, between billabongs and surrounding land, and from the grass understory into open water; this would simultaneously increase the variance in salvinia

**Table 1 | Estimates from the best-AIC fitting model of the biologically relevant parameters**

Parameter	Value	Description
$g_{\min}$	0.91	Minimum weevil-invulnerable tissue
$g_{\max}$	0.94	Maximum weevil-invulnerable tissue
$b_1$	5.41	Salvinia morphology inflection point
$b_2$	385.7	Slope of morphology change at $b_1$
$a$	0.09	Weevil attack rate
$k$	3.38	Aggregation parameter for weevils
$c$	0.94	Weevil reproduction scale parameter
$m$	0.0085	Salvinia net immigration
$v$	0.79*	Salvinia self-regulation form parameter
$d_s$	0†	Change in salvinia with water flow, $\delta_s(z_t) = d_s z_t$
$d_w$	−0.24	Change in weevils with water flow, $\delta_w(z_t) = d_w z_t$
$\sigma_{s1}$	0.21	Salvinia process error, $\text{var}\{e_s(z_t)\} = (\sigma_{s1} \exp(\sigma_{s2} z_t))^{\dagger} \tau_t$
$\sigma_{s2}$	0.35	Water flow effect on salvinia process error
$\sigma_{w1}$	0.42	Weevil process error, $\text{var}\{e_w(z_t)\} = (\sigma_{w1} \exp(\sigma_{w2} z_t))^{\dagger} \tau_t$
$\sigma_{w2}$	−0.08‡	Water flow effect on weevil process error

AIC, Akaike information criterion. The model without alternative stable states ( $g_{\min} = g_{\max}$ ,  $b_1 = b_2 = 0$ ) has a  $\Delta\text{AIC}$  of 48.76 and provides a statistically significantly inferior fit to the data ( $\chi^2 = 52.76$ ,  $P < 0.001$ ). For the model,  $R^2 = 0.73$  and  $0.38$  for log salvinia biomass and logit weevil damage, respectively. All values are statistically significant by likelihood ratio tests except as follows. \*Not different from 1 ( $\Delta\text{AIC} = 1.08$ ). †Not different from 0.01 ( $\Delta\text{AIC} = 1.82$ ). ‡Not different from 0 ( $\Delta\text{AIC} = 1.86$ ).



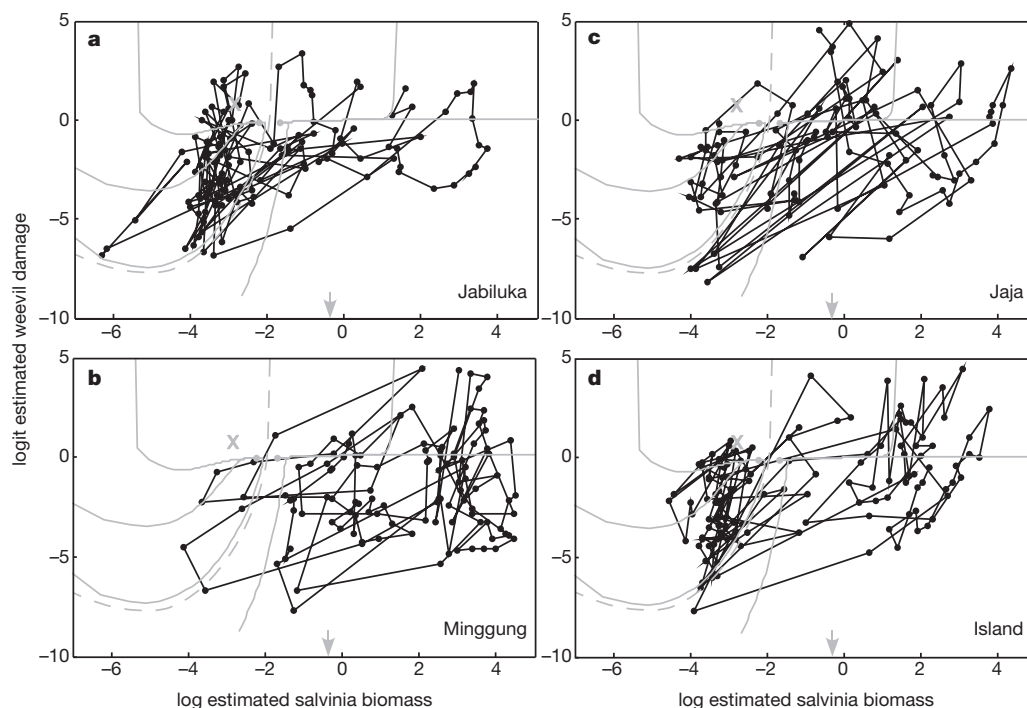
biomass and cause a net reduction of weevil abundance as salvinia colonizes billabongs from weevil-free refuges.

Existence of the alternative states requires the proportion of salvinia biomass in the category vulnerable to weevil attack to change with salvinia biomass; if the model is constrained so that this proportion does not change ( $g_{\min} = g_{\max}$ ), then alternative stable states are impossible, and the fit of the model is statistically significantly reduced (likelihood ratio test,  $\chi^2_3 = 52.76$ ,  $P \ll 0.001$ ). This provides strong support for the existence of alternative stable states. In three of the billabongs, the system spends time in the domains of attraction to both stable states, whereas Minggung has generally high salvinia abundance and rarely occupies the domain of weevil control (Fig. 3). We fit the model assuming that the underlying processes were identical across all billabongs. The poor salvinia control in Minggung could be due simply to the stochastic nature of the dynamics, by chance never staying long in the region of weevil control. There may be other differences between Minggung and other billabongs that we cannot identify, although such differences are not required to explain the observed dynamics.

Owing to the annual flooding events and high stochasticity in the system, the fit of the model relies not only on the existence of alternative stable states, but also the transient dynamics of the corresponding deterministic model. Although the deterministic model gives alternative stable states when the logarithm of water flow is fixed at its mean value, when the log water flow is fixed at one standard deviation below its mean, there is only a single stable point, and when fixed at one standard deviation above its mean, the weevil is eliminated from the system (Supplementary Information). Therefore, as the water flow regime fluctuates through its annual cycle, the alternative stable points alternately disappear; a possible, although still incomplete, description is that there are two alternative, environmentally forced cycles that have separate domains of attraction (Supplementary Fig. 8). Nonetheless, the 'ghost' of the boundary between stable points still affects the transient dynamics of the system<sup>6,23,24</sup>. An added

complexity is that the dynamical forces around the two alternative stable states differ. In the domain of attraction to the state with high salvinia biomass, changes in biomass are slow compared to changes in weevil damage, as illustrated by the deterministic trajectories of the model (Fig. 3). This allows salvinia biomass to dynamically wander between high and moderate values. The domain of attraction to the lower stable state contains trajectories that tend to approach the stable point through increases in both salvinia biomass and weevil damage, causing a positive correlation between these two variables. All of these dynamical patterns contribute to the strength of fit of the model to the data.

Our analyses point to possible opportunities to foster biological control of salvinia at Kakadu. For many examples of ecological systems with alternative stable states, the stability of the states makes transitions between them ecologically difficult and operationally challenging from a management perspective<sup>8,12</sup>. The Kakadu billabongs, however, are highly stochastic and experience periodic flooding, giving a window of opportunity to shift the system between states<sup>25,26</sup>. Even for Minggung, where biological control has been ineffective, the system occasionally jumps into the domain of attraction to the lower state of salvinia control. This occurs towards the end of the dry season after weevil populations recover from depression during flooding. If weevil control could be augmented at this time by inoculating Minggung with infested salvinia from other billabongs, the system could be captured in the lower domain of attraction. An alternative strategy would be to chemically or mechanically reduce salvinia as it is recovering from a flushing event, thereby allowing weevils more time to exert control. There is no guarantee that these strategies would work the first, the second, or even the third try. However, our theoretical demonstration that this lower state probably exists for Minggung should give hope for repeated attempts. Although alternative states are generally thought to present severe management challenges, when they are identified and understood, they may also present management solutions<sup>8,27</sup>.



**Figure 3 | Phase portraits of logit weevil damage against log salvinia biomass for the four billabongs.** a–d, Values have been standardized to have mean zero across all billabongs. Plotted trajectories (black dots and lines) are updated values fitted to the data, thereby smoothing the data to account for measurement error (Box 1). The dashed grey line divides the domains of attraction to the two stable points, marked with grey dots, and example deterministic trajectories are included (solid grey lines) when the drainage

water flow is fixed at its mean value, the time step between samples is assumed to be  $\tau = 8.83$  days, and there is no stochasticity in the model. The grey cross gives the equilibrium abundance of salvinia when there is no stochasticity and the log water flow is fixed at one standard deviation below its mean value. The grey arrow gives the deterministic abundance of salvinia in the absence of weevils which occurs when the flow rate is one standard deviation above its mean value (Supplementary Information).

## BOX 1:

## Salvinia–weevil model

We fit a single model to the data simultaneously for all four billabongs, thereby assuming that their dynamics are governed by the same processes. The model has a nonlinear state–space form<sup>28</sup>, with one set of equations describing the biological processes driving the dynamics and the other describing the sampling used to generate the data. Both process and measurement equations contain stochastic elements, with process error encapsulating environmental variation and measurement error describing any deviations between the ‘true’ state of the process variables and the data.

The process equations are:

$$x_{t+1} = \{ [x_t g(x_t) + x_t(1 - g(x_t))e^{r_t}(1 + (1 - g(x_t))x_t)^{-v}(1 - p_t)] \exp(\delta_s(z_t)) + m_{t1} \} \exp(\varepsilon_s(z_t))$$

$$y_{t+1} = \{ c x_t (1 - g(x_t))e^{r_t} p_t \exp(\delta_w(z_t)) \} \exp(\varepsilon_w(z_t))$$

where  $x_t$  and  $y_t$  are the salvinia biomass and weevil abundance at time  $t$ ,

$p_t = 1 - \left( 1 + \frac{a_{t1} y_t}{k(1 - g(x_t))x_t} \right)^{-k}$  is the proportion of susceptible

salvinia attacked by weevils assuming that the distribution of attacks is given by a negative binomial with aggregation parameter  $k$ , and  $\tau_t$  is the time interval between model iterations (averaging 8.83 days). The function  $g(x_t)$  gives the proportion of salvinia biomass in the category invulnerable to weevil damage.  $g(x_t)$  follows an inverse-logit function with minimum and maximum values  $g_{\min}$  and  $g_{\max}$ , inflection point where  $x_t = b_1$  and slope at the inflection point given by  $b_2$ . The input variable  $z_t$  is the log water flow measured at monitoring stations in each drainage, and  $\delta_s(z_t)$  and  $\delta_w(z_t)$  give the response of salvinia and weevil survival rates dependent on water flow.

The measurement equations are:

$$X_t^* = X_t + \log(g(e^{X_t})) + C + \alpha_s$$

$$W_t^* = \text{logit}(p_t) + \alpha_w$$

where  $X_t = \log x_t$  is the ‘true’ log salvinia biomass from the process equations,  $X_t^*$  is the observed log salvinia biomass assuming that only the invulnerable category is sufficiently dense to occur in visual sampling, and  $W_t^*$  is the logit of the observed weevil damage. The constant  $C$  is an overall scaling term for salvinia biomass, because the process equations are non-dimensional. The random variables  $\alpha_s$  and  $\alpha_w$  give measurement error, with  $\alpha_s$  assumed to have a Gaussian distribution with mean 0 and variance  $\sigma_{ms}^2$ , and  $\alpha_w$  assumed to have a quasi-binomial distribution with variance  $\left( \frac{1}{n} + \sigma_{mw}^2 \right) \frac{1}{p_t(1 - p_t)}$ ; if  $\sigma_{mw}^2 = 0$ , this would be the variance under a binomial distribution with a sample size of  $n$ , but to allow greater-than-binomial distribution, we also estimated  $\sigma_{mw}^2$ .

The model was fit using an extended Kalman filter to estimate the likelihood function<sup>28</sup>. The Kalman filter is an iterative algorithm in which the ‘true’ population sizes and estimates of their uncertainties are projected forward using the model. When an iteration coincides with a sample point, the true population sizes are updated using the observed values and the estimates of the measurement error; if the measurement error is small, then updating pulls values closer to their observed values. The model values plotted in Figs 2 and 3 are these updated values. The parameter  $r$ , the intrinsic rate of increase of the salvinia, was set at 0.08 per day as determined by extensive experiments at Kakadu (M.H.J., unpublished data). Backwards model selection was performed to find the best-fitting model, and the best-fitting model was confirmed with forward selection. Likelihood ratio tests were used to assess the statistical significance of key variables. Model parameters are described in Table 1, and detailed descriptions of the model and fitting procedure are given in the Supplementary Information.

Received 30 August; accepted 2 December 2010.

- Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
- Holling, C. S. Resilience and stability of ecological systems. *Annu. Rev. Ecol. Syst.* **4**, 1–23 (1973).
- May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* **269**, 471–477 (1977).
- Sutherland, J. P. Multiple stable points in natural communities. *Am. Nat.* **108**, 859–873 (1974).
- Folke, C. *et al.* Regime shifts, resilience, and biodiversity in ecosystem management. *Annu. Rev. Ecol. Syst.* **35**, 557–581 (2004).
- Van Geest, G. J., Coops, H., Scheffer, M. & van Nes, E. H. Long transients near the ghost of a stable state in eutrophic shallow lakes with fluctuating water levels. *Ecosystems* **10**, 36–46 (2007).
- Scheffer, M. & Carpenter, S. R. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends Ecol. Evol.* **18**, 648–656 (2003).
- Scheffer, M. *Critical Transitions in Nature and Society* (Princeton Univ. Press, 2009).
- Schröder, A., Persson, L. & De Roos, A. M. Direct experimental evidence for alternative stable states: a review. *Oikos* **110**, 3–19 (2005).
- Scheffer, M. *et al.* Early-warning signals for critical transitions. *Nature* **461**, 53–59 (2009).
- Carpenter, S. R., Brock, W. A., Cole, J. J., Kitchell, J. F. & Pace, M. L. Leading indicators of trophic cascades. *Ecol. Lett.* **11**, 128–138 (2008).
- Suding, K. N., Gross, K. L. & Houseman, G. R. Alternative states and positive feedbacks in restoration ecology. *Trends Ecol. Evol.* **19**, 46–53 (2004).
- Beisner, B. E., Haydon, D. T. & Cuddington, K. Alternative stable states in ecology. *Front. Ecol. Environ.* **1**, 376–382 (2003).
- Room, P. M. Ecology of a simple plant–herbivore system: biological control of *Salvinia*. *Trends Ecol. Evol.* **5**, 74–79 (1990).
- Room, P. M. & Thomas, P. A. Population growth of the floating weed *Salvinia molesta*: field observations and a global model based on temperature and nitrogen. *J. Appl. Ecol.* **23**, 1013–1028 (1986).
- Finlayson, C. M. Growth rates of *Salvinia molesta* in Lake Moondarra, Mount Isa, Australia. *Aquat. Bot.* **18**, 257–262 (1984).
- Julien, M. H., Hill, M. P. & Tipping, P. W. In *Biological Control of Weeds* (eds Muniappan, R., Reddy, G. V. P. & Raman, A.) 378–407 (Cambridge Univ. Press, 2009).
- Storrs, M. J. & Julien, M. H. In *Northern Landscapes Occasional Papers Vol. 1* (Australian Nature Conservation Agency, 1996).
- Room, P. M. & Thomas, P. A. Nitrogen, phosphorus and potassium in *Salvinia molesta* Mitchell in the field: effects of weather, insect damage, fertilizers and age. *Aquat. Biol.* **24**, 213–232 (1986).
- Scheffer, M., van Nes, E. H., Holmgren, M. & Hughes, T. Pulse-driven loss of top-down control: the critical-rate hypothesis. *Ecosystems* **11**, 226–237 (2008).
- Knowlton, N. Thresholds and multiple stable states in coral-reef community dynamics. *Am. Zool.* **32**, 674–682 (1992).
- Schmitz, O. J., Kalies, E. L. & Booth, M. G. Alternative dynamic regimes and trophic control of plant succession. *Ecosystems* **9**, 659–672 (2006).
- Hastings, A. Transients: the key to long-term ecological understanding? *Trends Ecol. Evol.* **19**, 39–45 (2004).
- Strogatz, S. H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (Perseus, 1994).
- Holmgren, M. & Scheffer, M. El Niño as a window of opportunity for the restoration of degraded arid ecosystems. *Ecosystems* **4**, 151–159 (2001).
- Holmgren, M. *et al.* Extreme climatic events shape arid and semiarid ecosystems. *Front. Ecol. Environ.* **4**, 87–95 (2006).
- Firn, J., House, A. P. N. & Buckley, Y. M. Alternative states models provide an effective framework for invasive species control and restoration of native communities. *J. Appl. Ecol.* **47**, 96–105 (2010).
- Harvey, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge Univ. Press, 1989).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Winderlich (Kakadu National Park) for facilitating data processing, and S. Cruickshank and G. Willis of the Northern Territory Department of Natural Resources, Environment, the Arts and Sport for providing gauging station data. M. Storrs, C. Murakami, F. Hunter, M. Hatt and F. Baird provided field assistance. A. Bourne assisted with data interpretation, and P. Milewski provided value insight into the mathematical analyses. S.S.S. received funding from the Australian Department of Agriculture, Fisheries and Forestry and A.R.I. received funding from a CSIRO McMaster’s Fellowship and United States National Science Foundation funds through the North Temperate Lakes Long Term Ecological Research Network and individual grants. S. R. Carpenter, Y. Buckley, and R. van Klinken provided comments on earlier versions of this manuscript.

**Author Contributions** M.H.J. developed the initial hypothesis explaining the partial success of salvinia biological control and designed the CSIRO sampling program at Kakadu. B.S. managed the Kakadu sampling program, and S.S.S. and B.S. assembled the data set. S.S.S. and A.R.I. designed the overall concept of the project reported here, and A.R.I. performed the analyses.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.R.I. (arives@wisc.edu).

# A thymus candidate in lampreys

Baubak Bajoghli<sup>1</sup>, Peng Guo<sup>2,3</sup>, Narges Aghaallaei<sup>1</sup>, Masayuki Hirano<sup>2,3</sup>, Christine Strohmeier<sup>1</sup>, Nathanael McCurley<sup>2,3</sup>, Dale E. Bockman<sup>4</sup>, Michael Schorpp<sup>1</sup>, Max D. Cooper<sup>2,3</sup> & Thomas Boehm<sup>1</sup>

Immunologists and evolutionary biologists have been debating the nature of the immune system of jawless vertebrates—lampreys and hagfish—since the nineteenth century. In the past 50 years, these fish were shown to have antibody-like responses and the capacity to reject allografts<sup>1</sup> but were found to lack the immunoglobulin-based adaptive immune system of jawed vertebrates<sup>2</sup>. Recent work has shown that lampreys have lymphocytes that instead express somatically diversified antigen receptors that contain leucine-rich-repeats, termed variable lymphocyte receptors (VLRs)<sup>3,4</sup>, and that the type of VLR expressed is specific to the lymphocyte lineage: T-like lymphocytes express type A VLR (*VLRA*) genes, and B-like lymphocytes express *VLRB* genes<sup>5</sup>. These clonally diverse anticipatory antigen receptors are assembled from incomplete genomic fragments by gene conversion<sup>6–9</sup>, which is thought to be initiated by either of two genes encoding cytosine deaminase<sup>9</sup>, cytosine deaminase 1 (*CDA1*) in T-like cells and *CDA2* in B-like cells<sup>5</sup>. It is unknown whether jawless fish, like jawed vertebrates, have dedicated primary lymphoid organs, such as the thymus, where the development and selection of lymphocytes takes place<sup>10,11</sup>. Here we identify discrete thymus-like lympho-epithelial structures, termed thymoids, in the tips of the gill filaments and the neighbouring secondary lamellae (both within the gill basket) of lamprey larvae. Only in the thymoids was expression of the orthologue of the gene encoding forkhead box N1 (FOXN1)<sup>10</sup>, a marker of the thymopoietic microenvironment in jawed vertebrates<sup>12</sup>, accompanied by expression of *CDA1* and *VLRA*. This expression pattern was unaffected by immunization of lampreys or by stimulation with a T-cell mitogen. Non-functional *VLRA* gene assemblies were found frequently in the thymoids but not elsewhere, further implicating the thymoid as the site of development of T-like cells in lampreys. These findings suggest that the similarities underlying the dual nature of the adaptive immune systems in the two sister groups of vertebrates extend to primary lymphoid organs.

The long-standing question of whether lampreys have a thymus led to the description of several circumscribed lymphoid accumulations in the gill basket of lamprey larvae as candidates for the thymus, for instance in the lateral branchial wall sinuses<sup>13</sup> and the ‘thymic placode’ inside the epipharyngeal folds<sup>14</sup> (Supplementary Information and Supplementary Fig. 1). However, histological surveys failed to reveal a thymus analogue unambiguously. We were prompted to re-examine the issue of a thymus equivalent in lampreys by the recent identification of two separate lineages of lymphocytes in lampreys, a finding indicating that the dual nature of the immune system extends to all vertebrates<sup>5</sup>. T-like cells of lampreys were shown to express *VLRA* on their surface (denoted *VLRA*<sup>+</sup> cells), whereas B-like cells express surface and secreted forms of *VLRB*. Moreover, *VLRA*<sup>+</sup> cells were shown to express *CDA1* preferentially, whereas *VLRB*<sup>+</sup> cells express *CDA2* (ref. 5). With respect to a putative thymus primordium in jawless fish, the lamprey orthologue of *FOXN1*, a marker of the thymopoietic microenvironment in jawed vertebrates<sup>11,12</sup>, was found to be expressed in lamprey larvae in a region of the gill basket where *VLRB*-expressing

lymphocytes could not be found<sup>10</sup>. On the basis of the cell-type specificity of gene expression in lamprey lymphocytes, we therefore hypothesized that co-expression of *VLRA* and *CDA1* or of *VLRB* and *CDA2* would distinguish the primary lymphoid organs in which these T-like and B-like cells are generated in jawless vertebrates.

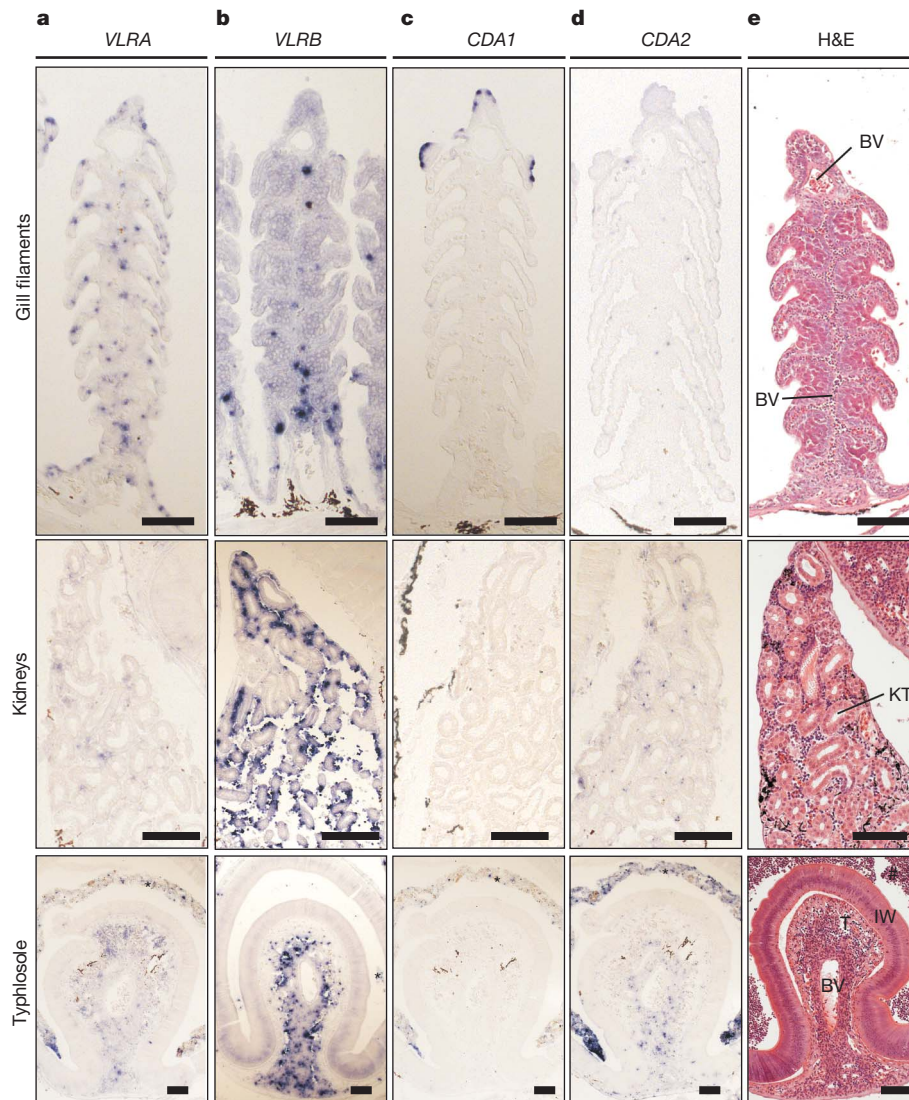
Expression of the *VLR* genes was detectable by RNA *in situ* hybridization in cells of many tissues of European brook lamprey (*Lampetra planeri*) larvae. Cells expressing *VLRA* and *VLRB* messenger RNAs were found in the gill filaments, kidneys, typhlosole and blood (Fig. 1a, b). *VLRA*-expressing cells dominated in the gill basket, whereas *VLRB*-expressing cells were more frequent in the kidneys and typhlosole (Supplementary Fig. 2). These data are in agreement with previous flow cytometric analyses of cells from the closely related sea lamprey (*Petromyzon marinus*)<sup>5</sup> and indicate a tissue-specific pattern of distribution for the two lymphocyte lineages.

Expression of the *CDA* genes was more spatially restricted than that of the *VLR* genes. When assayed by *in situ* hybridization, *CDA1* mRNA was detected at discrete sites that were located primarily at the tip of gill filaments. *CDA1* was not expressed by cells in the liver (not shown), kidneys, typhlosole, blood or other tissues outside the gill basket (Fig. 1c). Conversely, cells expressing *CDA2* were preferentially expressed in the typhlosole, predominantly around the central blood vessel, and occasionally expressed in the kidneys and blood (Fig. 1d), indicating that the sites of *CDA1* and *CDA2* expression do not overlap. The patterns of co-expression of *VLR* and *CDA* genes in these distinct anatomical sites (Fig. 1e) suggest that *VLRA*<sup>+</sup> T-like cells might develop in the gill region, whereas *VLRB*<sup>+</sup> B-like cells might originate in haematopoietic tissue, namely in the typhlosole and kidneys. This is reminiscent of the situation in jawed vertebrates, in which T cells develop in the thymus and B cells develop primarily in the bone marrow or its functional equivalents.

The thymus in jawed vertebrates is uniquely identifiable by thymocyte expression of recombination-activating genes (RAGs) during rearrangement of the genes that encode the T-cell antigen receptor, as well as by epithelial cell expression of the gene encoding the transcription factor FOXN1. It has previously been shown that the lamprey orthologue of the thymic epithelial marker FOXN1, designated FOXN1 (or FOXN4L), is expressed in the epithelium of the gill basket of lamprey larvae<sup>10</sup>. To explore the possibility that a thymus-like structure might be found in the gill basket of lamprey larvae, we searched for sites at which markers of thymic epithelial cells (*FOXN1* and Delta-like B (*DLL-B*)) and of lymphocytes (*VLRA* and *CDA1*) were expressed together. All four genes were found to be expressed in the same region of the gill filaments (Fig. 2a). Double *in situ* hybridization confirmed the association of *VLRA* and *CDA1* expression in the thymoids and indicated that some *VLRA*-expressing cells also express *CDA1*. By contrast, *VLRB*-expressing cells do not express *CDA1* (Fig. 2b) and are absent from the thymoid region (Figs 1b and 2b). To examine the micro-anatomical relationship between epithelial and lymphocyte markers, we used a *DLL-B* probe that specifically detects the expression of a putative FOXN1 target gene encoding a Delta-related Notch

<sup>1</sup>Department of Developmental Immunology, Max Planck Institute of Immunobiology and Epigenetics, Stuebeweg 51, D-79108 Freiburg, Germany. <sup>2</sup>Emory Vaccine Center, 954 Gatewood Road, Atlanta, Georgia 30329, USA. <sup>3</sup>Department of Pathology and Laboratory Medicine, Emory University, 1462 Clifton Road North-East, Atlanta, Georgia 30322, USA. <sup>4</sup>Department of Cellular Biology and Anatomy, Medical College of Georgia, Augusta, Georgia 30912, USA.





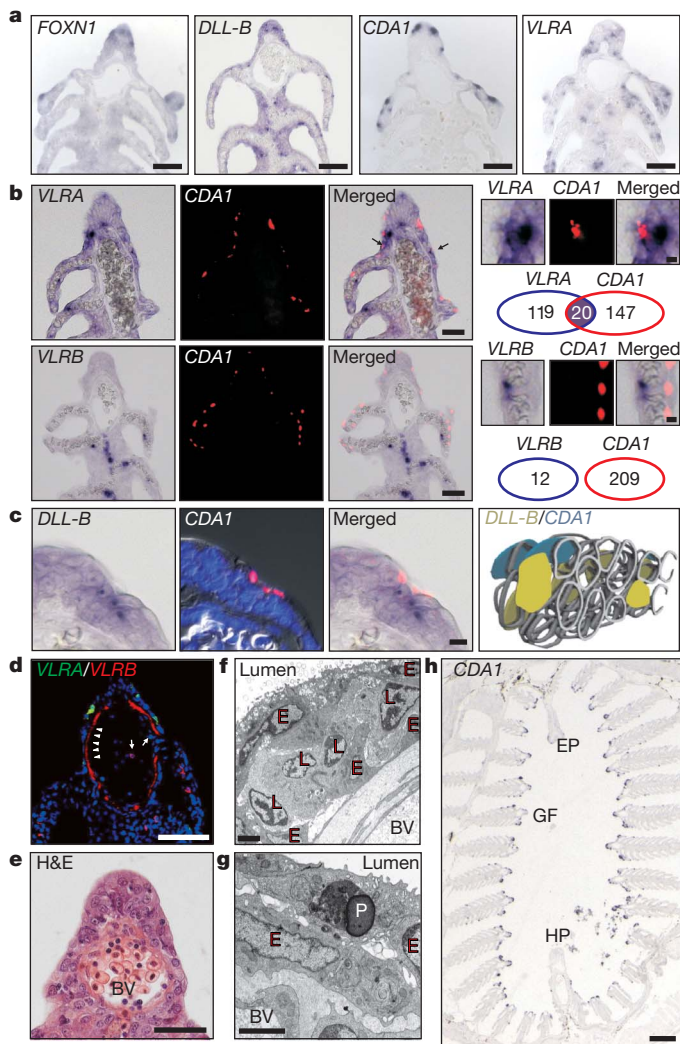
**Figure 1 | Tissue-specific expression of VLR and CDA genes in *L. planeri* larvae.** **a–d**, Gene-specific expression, as determined by RNA *in situ* hybridization with gene-specific riboprobes, is shown in blue. **a**, VLRA expression. **b**, VLRB expression. **c**, CDA1 expression. **d**, CDA2 expression. **e**, Histological structure. Haematoxylin and eosin (H&E) staining of sections

equivalent to those used for RNA *in situ* hybridization. The intestinal contents (\*) (primarily yeast) cause non-specific staining in **a–d** and are also seen in **e** (#). BV, blood vessel; IW, intestinal wall; KT, kidney tubule; T, typhlosole. Scale bars, 100  $\mu$ m.

ligand<sup>10</sup>, as a marker of epithelial cells, and CDA1, as a marker of developing T-like cells. In the thymoid, these cells are located in close proximity (Fig. 2c). Analysis by immunofluorescence confirmed the presence of VLRA-producing cells and the absence of VLRB-producing cells in the thymoid, whereas both cell types were found in blood vessels (Fig. 2d). Histological analysis by light microscopy showed that the thymoid contains both epithelial cells and lymphocytes (Fig. 2e). Electron micrographs confirmed that these cells are present in close proximity in the thymoid (Fig. 2f) and indicate that phagocytosis of cellular debris occurs at this site (Fig. 2g). Only a small subset of CDA1-expressing cells proliferate (Supplementary Fig. 3), suggesting cell-cycle-specific regulation of this lamprey gene belonging to the AID–APOBEC family. Cells expressing CDA1 were found in the tips of the gill filaments encompassing the entire gill basket (Fig. 2h and Supplementary Fig. 4), indicating that the thymoids are not confined to a specific pharyngeal arch. These findings in *L. planeri* were confirmed for *P. marinus* (Supplementary Fig. 5), indicating that the specific gene expression patterns are a general characteristic of

lampreys. We conclude that the lympho-epithelial structure identified in the gill basket bears the diagnostic hallmarks expected for a thymus equivalent: namely expression of CDA1 (a gene that is predicted to be associated with somatic diversification of the VLRA locus) together with VLRA expression, and the concordant expression of FOXP1 with DLL-B (one of the key target genes of FOXP1).

In jawed vertebrates, the thymus is a primary lymphoid organ, which, unlike secondary lymphoid tissues such as lymph nodes and the spleen, does not change its structure during an immune response. We applied this criterion to the thymoid. Lamprey larvae mount strong immune responses against various antigens, including the exo-sporium of *Bacillus anthracis*<sup>5,15</sup>. Immunization with this antigen (Fig. 3a) leads to a general proliferative response in haematopoietic tissues<sup>15</sup>. There is a massive increase in cell proliferation in the typhlosole and kidneys, whereas the increase in the number of proliferating cells in the gill vasculature is more modest (Fig. 3b). Concomitantly, the number of VLRA-expressing cells increases in the typhlosole, and the number of VLRB-expressing cells increases in the blood vessels,



**Figure 2 | Characterization of the lamprey thymoid.** **a**, Concomitant expression of epithelial markers (*FOXN1* and *DLL-B*) and two lymphocyte-specific genes (*CDA1* and *VLRA*) at the tip of gill filaments, as determined by RNA *in situ* hybridization with gene-specific riboprobes. Scale bars, 50  $\mu$ m. **b**, Co-expression of *VLRA* (blue) and *CDA1* (red) in a subset of *VLRA*-expressing lymphocytes (upper panels), as determined by RNA *in situ* hybridization with gene-specific riboprobes on adjacent sections, which are shown separately and as a merged image. Co-expression seems to occur preferentially in cells expressing low levels of *VLRA* (arrows indicate co-expressing cells). *VLRB* (blue) and *CDA1* are not co-expressed (lower panels). Scale bars, 40  $\mu$ m. Greater magnifications are shown to the right (scale bars, 5  $\mu$ m) with quantitative assessments of gene expression presented underneath (numbers indicate number of cells that express one or both genes). **c**, Close proximity of cells expressing *DLL-B* (blue) and *CDA1* (red) in the thymoid. RNA *in situ* hybridization was carried out with gene-specific locked-nucleic-acid-enhanced oligonucleotides on adjacent sections, which are shown separately and as a merged image. For the *CDA1* panel, the images of the DNA counterstain (4',6-diamidino-2-phenylindole (DAPI), blue) and the differential interference contrast are superimposed. Scale bar, 10  $\mu$ m. The rightmost panel shows a three-dimensional reconstruction of thymoid tissue with *DLL-B*- and *CDA1*-expressing cells indicated in different colours; other cells are shown only in outline. **d**, Immunofluorescence microscopy with anti-*VLRA* (green) and anti-*VLRB* (red) antibodies. Intravascular lymphocytes are indicated by arrows. Deposits of *VLRB* at the vessel walls are indicated by arrowheads. DNA is counterstained with DAPI (blue). Scale bar, 50  $\mu$ m. **e**, Lympho-epithelial structure of the thymoid as revealed by light microscopy. Nucleated erythrocytes are visible in the large blood vessel (BV) underneath the thymoid. Scale bar, 30  $\mu$ m. **f**, Lympho-epithelial structure of the thymoid as revealed by transmission electron microscopy: lymphocytes (L), epithelial cells (E) and blood vessel (BV). The lumen of the gill chamber is at the top left. Scale bar, 2  $\mu$ m. **g**, Evidence for phagocytosis (P) in the thymoid. Scale bar, 2  $\mu$ m. **h**, *CDA1* expression is found at the tips of gill filaments (GF) and in the epithelial linings of the epipharyngeal ridge (EP) and the hypopharyngeal fold (HP), as shown by RNA *in situ* hybridization with a gene-specific riboprobe on a transverse section (see also Supplementary Fig. 4). Scale bar, 300  $\mu$ m. Sections are from *L. planeri* in **a** (*FOXN1*, *CDA1* and *VLRA*), **e** and **h**, and from *P. marinus* in all other panels.

being particularly noticeable in the cavernous bodies located at the base of gill filaments (Supplementary Fig. 6). After intracoelomic injection of phytohaemagglutinin (PHA) (Fig. 3a), this mitogen is distributed throughout the vasculature and rapidly accumulates in the gill region (Supplementary Fig. 7). PHA is known to cause a proliferative burst of *VLRA*<sup>+</sup> cells<sup>5</sup>, and this general proliferative effect is clearly seen by using *in situ* analysis in the gill region, kidneys and typhlosole (Fig. 3b). Stimulation with PHA also increased the number of *VLRA*-expressing cells in the gill region, kidneys and typhlosole (Supplementary Fig. 6). Importantly, however, PHA does not induce proliferation of cells in the thymoid tips (Fig. 3b), and it does not change the pattern of *CDA1* expression (Fig. 3c). The finding that cells in the thymoids fail to respond to antigenic or mitogenic stimuli is compatible with the thymoid being a primary lymphoid organ rather than a secondary one.

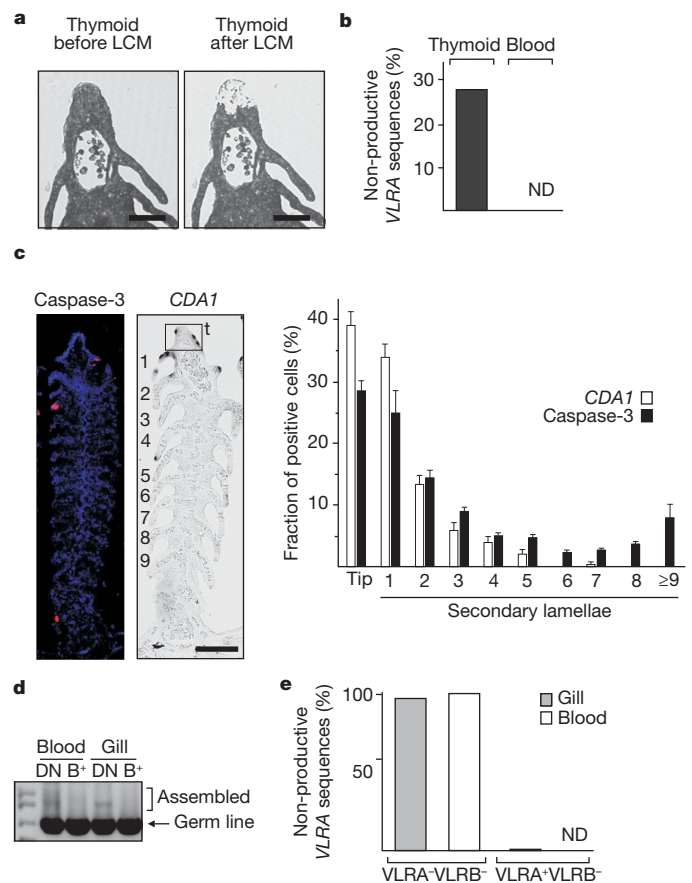
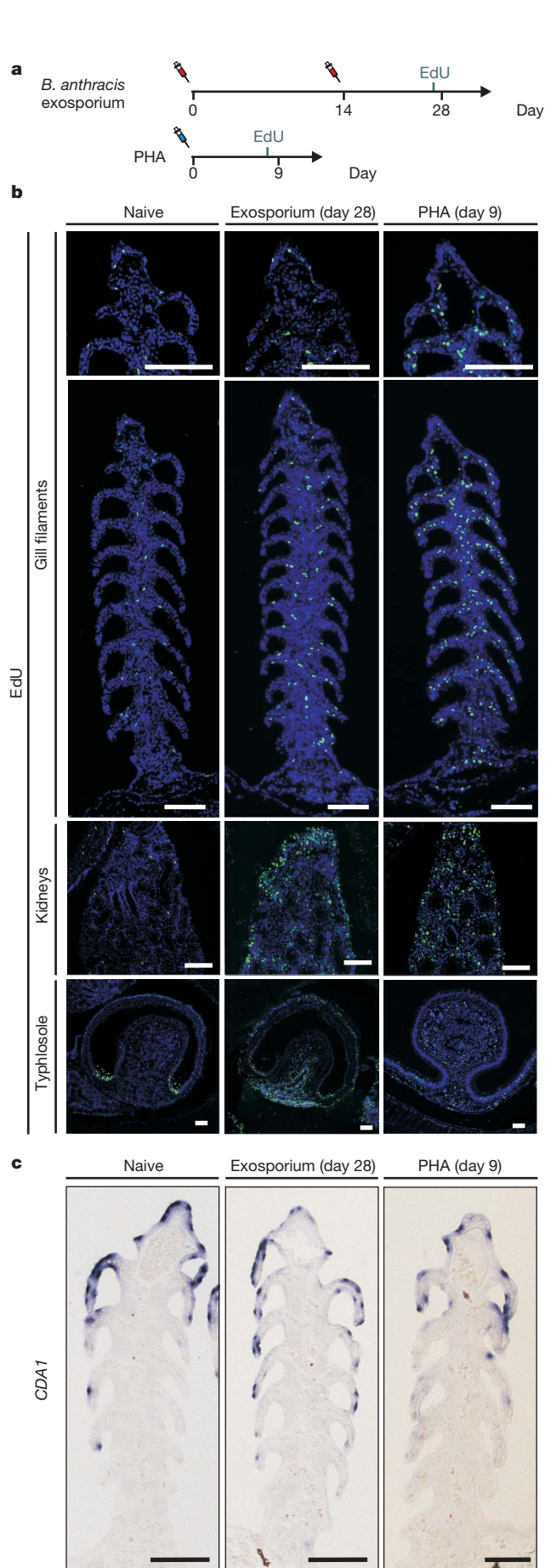
We sought more direct evidence that the thymoid is the site at which T-like cells develop. We examined the status of *VLRA* gene assembly in this region and elsewhere. Previously, it has been shown that *VLRA*<sup>+</sup> lymphocytes typically assemble only one *VLRA* allele<sup>5</sup>, the other allele being retained in the germ-line configuration. Furthermore, the assembled *VLRA* allele is almost always functional. Non-functional rearrangements of *VLRA* genes are exceedingly rare and are always accompanied by a functional assembly on the second allele<sup>16</sup>. This observation suggests a process that ensures the selective development of *VLRA*<sup>+</sup> lymphocytes. We used laser-capture microdissection to procure genomic DNA from cells at the thymoid tip of the gill filament and, as a control, from within the blood vessels at the filament base (Fig. 4a). As anticipated, assembled *VLRA* genes could be amplified from the DNA isolated from both the thymoid and the blood.

Strikingly, however, non-functional sequences were found only in the thymoid fraction, whereas the sequences obtained from cells isolated from blood vessels were all functional (Fig. 4b and Supplementary Fig. 8). This difference indicates that non-functional assemblies impair the further development of *VLRA*<sup>+</sup> lymphocytes. Indeed, within the gill basket, caspase-3-positive apoptotic cells were also found primarily in the gill filament tip and neighbouring secondary lamellae, closely resembling the distribution of *CDA1*-expressing cells (Fig. 4c). These findings are compatible with the idea of selection against thymoid cells that are undergoing non-functional *VLRA* assembly. It will be interesting to examine whether this is also true for *VLRB*, which was described recently and is structurally similar to *VLRA*<sup>17</sup>.

To evaluate the stringency of this selection process, we investigated whether non-functionally assembled *VLRA* genes could be identified in *VLRA*<sup>+</sup>*VLRB*<sup>+</sup> cells isolated from the blood and gills. Although it was possible to amplify assembled *VLRA* genes from *VLRA*<sup>+</sup>*VLRB*<sup>+</sup> cells with lymphocyte light-scattering features, two rounds of amplification were required to detect them (Fig. 4d). Essentially all of these relatively rare *VLRA* assemblies were non-functional, in contrast to the status of *VLRA* genes in *VLRA*<sup>+</sup>*VLRB*<sup>+</sup> cells (Fig. 4e and Supplementary Fig. 9). Notably, *VLRA* assemblies were not found in *VLRA*<sup>+</sup>*VLRB*<sup>+</sup> cells (Fig. 4d). Together, these findings indicate a remarkable efficiency of selection for *VLRA*<sup>+</sup> cells.

Here we describe a previously unrecognized lympho-epithelial structure in the gill basket of lamprey larvae. The thymoid bears the hallmarks of a tissue where lymphocytes undergo *VLRA* assembly and selection for the expression of functional *VLRA* genes, which encode the anticipatory antigen receptor of T-like lymphocytes. Our findings suggest that this structure is a candidate for a thymus in lampreys. The dispersed nature and relatively inconspicuous morphological appearance of the lamprey thymoid structure could explain why it has gone unnoticed for so long





**Figure 4 | Selection of VLRA-expressing lymphocytes in the thymoid.**

**a**, Micrographs showing the site of tissue procurement, by laser-capture microdissection (LCM), at the gill filament tip region of the thymoid. Scale bars, 50  $\mu$ m. **b**, Proportion of non-productive sequences among assembled VLRA genes. The difference between the thymoid and blood is significant at  $P < 0.01$  (Fisher's exact test) (thymoid,  $n = 29$ ; blood,  $n = 18$ ;  $n$ , number of sequences). ND, not detectable. **c**, Apoptotic cells (left), as detected with anti-caspase-3 antibody (red), preferentially occur in the thymoid, as detected by CDA1 expression (centre, blue). In the left panel, DNA is counterstained with DAPI (blue). In the centre panel, anatomical landmarks are indicated: tip of gill filament (t), first secondary lamella (1), second secondary lamella (2) and so on. A numerical analysis of the distributions of caspase-3-positive and CDA1-expressing cells relative to the relevant anatomical landmarks of gill filaments is shown in the right panel; randomly selected sections were used (caspase-3,  $n = 36$ ; CDA1,  $n = 17$ ;  $n$ , number of sections; error bars, s.e.m.). Scale bars, 100  $\mu$ m. **d**, Assembled VLRA genes are detectable in VLRA<sup>-</sup>VLRB<sup>-</sup> (double negative, DN) cells separated by FACS after isolation from the blood or gills of lamprey larvae but not in VLRA<sup>-</sup>VLRB<sup>+</sup> (B<sup>+</sup>) cells. **e**, Non-productive assemblies of VLRA genes predominate in VLRA<sup>-</sup>VLRB<sup>-</sup> cells. By contrast, in VLRA<sup>+</sup>VLRB<sup>-</sup> cells, essentially all assembled VLRA genes are productive. VLRA<sup>-</sup>VLRB<sup>-</sup> cells,  $n = 42$  for gills,  $n = 33$  for blood; VLRA<sup>+</sup>VLRB<sup>-</sup> cells,  $n = 33$  for gills,  $n = 39$  for blood;  $n$ , number of sequences. Samples are from *L. planeri* in **a** and **b** and *P. marinus* in **c–e**.

**Figure 3 | Stimulation does not affect cell proliferation and gene expression in the thymoid of *P. marinus*.** **a**, Time points for injection with *B. anthracis* exosporium, the lectin PHA and the nucleoside analogue 5-ethynyl-2'-deoxyuridine (EdU). **b**, Representative sections revealing the number and location of proliferating cells (EdU, green), as determined by immunofluorescence microscopy. DNA is stained with Hoechst stain (blue). The tips of the gill filaments are shown at higher magnification in the upper panels. **c**, Distribution of CDA1-expressing cells in gill filaments, as determined by RNA *in situ* hybridization with a gene-specific riboprobe. Scale bars, 100  $\mu$ m.



and why it was revealed only by extensive gene expression analyses. The identification of the thymoid in lampreys provides a starting point for more detailed comparative studies between jawless and jawed vertebrates, guided by the wealth of information about thymopoiesis in jawed vertebrates. Finally, these findings suggest that a common vertebrate ancestor had not only T-like and B-like lymphocyte lineages but also anatomically distinct tissues in which these cells could develop in a spatially segregated manner.

## METHODS SUMMARY

*Lampetra planeri* larvae were sampled in the Freiburg area (Germany). *Petromyzon marinus* larvae were collected from tributaries to Lake Michigan (Michigan). RNA *in situ* hybridization analyses were performed on paraffin-embedded tissue sections. Specific antibodies were used to detect VLRA<sup>+</sup> and VLRB<sup>+</sup> lymphocytes *in situ*. Cells undergoing apoptosis were detected with caspase-3-specific antibodies, and proliferating cells were marked by incorporation of 5-ethynyl-2'-deoxyuridine and detected by Click-iT reaction. Preparative flow cytometry was used to separate lymphocyte populations according to their VLRA and VLRB expression. PCR with specific primers was used to amplify germ-line and assembled forms of VLRA genes for subsequent cloning and sequence analysis. Laser-capture microdissection of paraffin-embedded tissue sections was used to procure genomic DNA from specific regions. Transmission electron microscopy was used for high-resolution tissue analysis.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 May; accepted 8 November 2010.**

1. Finstad, J. & Good, R. A. The evolution of the immune response. III. Immunologic responses in the lamprey. *J. Exp. Med.* **120**, 1151–1168 (1964).
2. Mayer, W. E. *et al.* Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl Acad. Sci. USA* **99**, 14350–14355 (2002).
3. Pancer, Z. *et al.* Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* **430**, 174–180 (2004).
4. Pancer, Z. *et al.* Variable lymphocyte receptors in hagfish. *Proc. Natl Acad. Sci. USA* **102**, 9224–9229 (2005).
5. Guo, P. *et al.* Dual nature of the adaptive immune system in lampreys. *Nature* **459**, 796–801 (2009).
6. Alder, M. N. *et al.* Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* **310**, 1970–1973 (2005).
7. Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* **124**, 815–822 (2006).
8. Nagawa, F. *et al.* Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nature Immunol.* **8**, 206–213 (2007).
9. Rogozin, I. B. *et al.* Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nature Immunol.* **8**, 647–656 (2007).
10. Bajoghli, B. *et al.* Evolution of genetic networks underlying the emergence of thymopoiesis in vertebrates. *Cell* **138**, 186–197 (2009).
11. Boehm, T. & Bleul, C. C. The evolutionary history of lymphoid organs. *Nature Immunol.* **8**, 131–135 (2007).
12. Nehls, M. *et al.* Two genetically separable steps in the differentiation of thymic epithelium. *Science* **272**, 886–889 (1996).
13. Schaffer, J. Über die Thymusanlage bei *Petromyzon planeri*. Zweite vorläufige Mittheilung über den feineren Bau des Thymus. *Sber. K. Akad. Wiss.* **103**, 149–156 (1894).
14. Wallin, I. E. The relationships and histogenesis of thymus-like structures in ammocoetes. *Am. J. Anat.* **22**, 127–167 (1917).
15. Alder, M. N. *et al.* Antibody responses of variable lymphocyte receptors in the lamprey. *Nature Immunol.* **9**, 319–327 (2008).
16. Kishishita, N. *et al.* Regulation of antigen-receptor gene assembly in hagfish. *EMBO Rep.* **11**, 126–132 (2010).
17. Kasamatsu, J. *et al.* Identification of a third variable lymphocyte receptor in the lamprey. *Proc. Natl Acad. Sci. USA* **107**, 14304–14308 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank C. Happe and M. Held for technical help, C. L. Turnbough Jr for providing *B. anthracis* spores and exosporia, and the Emory University School of Medicine core facility for flow cytometry for cell-sorting services. This work was supported by the Max Planck Society, the Deutsche Forschungsgemeinschaft, the National Institutes of Health and the Georgia Research Alliance.

**Author Contributions** B.B., P.G., N.A., M.H., C.S., N.M., D.E.B., M.S., M.D.C. and T.B. designed the experiments, performed the research, analysed the data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.B. ([boehm@immunbio.mpg.de](mailto:boehm@immunbio.mpg.de)).

## METHODS

**Animals and immune stimulation.** *Lampetra planeri* (5–12 cm long and 1–3 years of age) were sampled in the Freiburg area (Germany) with the permission of the local authorities. *Petromyzon marinus* larvae (8–12 cm long and 2–4 years of age) were from Lake Michigan tributaries in Michigan (Lamprey Services). Fish were maintained in sand-lined aquaria at 18 °C and fed brewer's yeast. Larvae were immunized with *Bacillus anthracis* exosporium or stimulated with phytohaemagglutinin (PHA) (Sigma) as described previously<sup>4</sup>. Briefly, larvae were anaesthetized with 0.1 g l<sup>-1</sup> ethyl 3-aminobenzoate methanesulphonate (MS-222) (Sigma) and were given 60 µl intracoelomic injections of *B. anthracis* exosporia (10 µg) or PHA-L (leukoagglutinin; 25 µg) prepared in sterile 0.67× PBS buffer. Immunization with exosporia was carried out on day 0 and 14, and tissues were collected after animals were killed by administering 1 g l<sup>-1</sup> MS-222 on day 28. PHA was administered as a single injection, and animals were killed on day 9 after injection. For proliferation assays, lampreys were injected with 5-ethynyl-2'-deoxyuridine (EdU) (5 µg) (Invitrogen) in 60 µl 0.67× PBS and returned to aquaculture for 24 h before being killed.

**Flow cytometry and cell sorting.** Leukocytes were isolated from the blood and other tissues of *P. marinus* larvae before being separated by FACS as previously described<sup>5</sup>. Briefly, blood was collected in 0.67× PBS containing 30 mM EDTA, and buffy coats were prepared by centrifugation at 50g. Leukocytes were liberated from gills by mincing with frosted glass slides. Cells were stained with R110 anti-VLRA rabbit polyclonal serum and 4C4 anti-VLRB mouse monoclonal antibody, followed by R-phycoerythrin-conjugated goat anti-rabbit antibody and allophycocyanin-conjugated goat anti-mouse antibody (SouthernBiotech), and then were sorted into VLRA<sup>+</sup>, VLRB<sup>+</sup> and VLRA<sup>-</sup>VLRB<sup>-</sup> populations on a FACSAria II flow cytometer (Becton Dickinson).

**Genomic DNA and PCR.** Genomic DNA was extracted from sorted cell populations—VLRA<sup>+</sup>, VLRB<sup>+</sup> and VLRA<sup>-</sup>VLRB<sup>-</sup> cells in the lymphocyte gate of blood and gill samples from *P. marinus*—using a DNeasy kit (QIAGEN). First-round PCR was carried out with VLRA-F and VLRA-R primers (and Ex Taq polymerase, Takara). Reactions were amplified using the following: one cycle of 94 °C for 1 min; 35 cycles of 94 °C for 15 s, 56 °C for 20 s and 72 °C for 1 min; and one cycle of 3 min at 72 °C. Second-round (nested) PCR was performed with VLRA-F2 and VLRA-R2 primers. Reactions were amplified using the following: one cycle of 94 °C for 1 min; 35 cycles of 94 °C for 15 s, 57 °C for 20 s and 72 °C for 1 min; and one cycle of 3 min at 72 °C. PCR products were cloned into the pGEM-T vector (Promega). Paraffin-embedded sections (5–7 µm) of *L. planeri* were used for laser-capture microdissection as previously described<sup>18</sup>. Genomic DNA extraction was performed using a PicoPure DNA extraction kit (Arcturus). VLRA genes from *L. planeri* tissues were amplified by nested PCR, the first amplification using primers vlra\_1f and vlra\_2r, the second amplification using primers vlra\_3f and vlra\_4r. Reactions were amplified using the following: first amplification, one cycle of 96 °C for 2 min; 35 cycles of 96 °C for 15 s, 60 °C for 30 s and 72 °C for 2 min; and one cycle of 5 min at 72 °C; and second amplification, as above but for 25 cycles instead of 35. The assembled VLRA gene fragments were then cloned into the pGEM-T vector. Clones were sequenced using the ABI 3730xl DNA Analyzer (Applied Biosystems). Primer sequences are listed in Supplementary Table 1.

**Immunofluorescence microscopy.** For analysis of lymphocyte distribution in *P. marinus* larvae, dissected tissues were fixed for 12 h in 0.67× PBS containing 2% paraformaldehyde at 4 °C, cryopreserved in 30% sucrose, embedded in OCT compound (Tissue-Tek, Sakura) and sectioned at 7 µm on a cryostat (Thermo). Sections were permeabilized in PBS containing 10% goat serum, 0.5% saponin, 10 mM HEPES buffer and 10 mM glycine. They were stained for 1 h with R110 and 4C4 antibodies, followed by 1 h with the appropriate Alexa-Fluor-conjugated secondary antibodies (Invitrogen) before being mounted in ProLong Gold with 4',6-diamidino-2-phenylindole (DAPI) solution (Invitrogen). Fluorescence microscopy was performed with an Axiovert 200M microscope (Zeiss), equipped with a ×40 objective (numerical aperture, 0.6; ocular magnification, ×10).

For detection of apoptotic cells, frozen sections (10–15 µm) were fixed in 4% paraformaldehyde for 15 min, washed several times in PBS, and permeabilized with PBS containing 0.2% Triton X-100 detergent for 5 min, then washed again in PBS. To block non-specific binding sites, the sections were incubated for 1 h in 5% normal goat serum diluted in blocking solution (PBS containing 0.1% Tween 20 detergent). Sections were then incubated with rabbit anti-active caspase-3 (G7481, Promega; at a 1:250 dilution in blocking solution) for 16 h at 4 °C. After washing in PBS containing 0.1% Tween 20, a Cy3-conjugated donkey anti-rabbit antibody (1:500 in blocking solution) was applied for 30 min. Sections were washed several times in PBS and, after drying, were mounted in Fluoromount-G and 4',6-diamidino-2-phenylindole (DAPI). Micrographs were taken using an Axio Imager.Z1 and ApoTome microscope (Zeiss).

For analysis of cell proliferation, lampreys were embedded in OCT and cryosectioned. EdU detection was performed using the Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (Invitrogen). Briefly, frozen sections (10–15 µm) were fixed with 4% paraformaldehyde for 15 min at 20 °C and washed in PBS. Sections were permeabilized with PBS containing 0.5% Triton X-100 and then washed in PBS. Sections were then incubated with the EdU reaction cocktail for 30 min at 20 °C. After several washes in PBS, sections were incubated in Hoechst stain (1:2,000 in PBS) for 30 min. Sections were washed several times in PBS and, after drying, were mounted with Fluoromount-G. For EdU detection in paraffin-embedded tissue, sections were deparaffinized and rehydrated, and then permeabilized and processed as above. Images were processed with the program Photoshop (Adobe).

**Electron microscopy.** Killed *P. marinus* larvae were immersed in Karnovsky's fixative (paraformaldehyde and glutaraldehyde) for 2 h. Thin cross sections were cut from gill regions and prepared for transmission electron microscopic analysis as described previously<sup>19</sup>. Sections were examined with a JEM-1230 transmission electron microscope (JEOL). Images were recorded digitally using the UltraScan 4000 imaging system (Gatan).

**In situ hybridization analysis.** RNA *in situ* hybridization was performed with digoxigenin (DIG)-labelled RNA riboprobes as described previously<sup>10</sup>. Probe sequences are listed in Supplementary Table 2. For *DLL-B* and *CDA1* colocalization studies, *in situ* hybridization was carried out by use of short gene-specific locked-nucleic-acid (LNA)-enhanced oligonucleotides (Exiqon; see Supplementary Table 2) as described previously<sup>10</sup>, with the exception that the temperature of hybridization was 54 °C. Double *in situ* hybridization was carried out as follows. DIG- and fluorescein isothiocyanate (FITC)-labelled RNA antisense probes were hybridized to RNA in tissue sections simultaneously. The DIG-labelled probe was detected first, with an alkaline-phosphatase-conjugated anti-DIG antibody (1:2,000 dilution in maleic acid buffer (MAB); 100 mM maleic acid, pH 7.5, 150 mM NaCl, 2 mM Levamisol, 0.1% Tween 20 and 1% blocking reagent (Roche)). It was revealed by staining with BM Purple (Roche), according to the manufacturer's instructions. The sections were washed several times with TNT solution (10 mM Tris-Cl, pH 7.5, 500 mM NaCl and 0.1% Tween 20), and the FITC-labelled probe was detected by a peroxidase-conjugated anti-FITC antibody (1:500 dilution in MAB) and revealed by Cy5 fluorescence, using Tyramide Signal Amplification Plus Cy5 system (Perkin Elmer). For three-dimensional reconstructions, alternate (5 µm) sections were hybridized with gene-specific LNA-enhanced oligonucleotides. After the staining reactions, the thymoid regions of the gill filaments were photographed using an Imager Z1 microscope (Zeiss). A series of TIF files indicating the positions of the *CDA1*- and *DLL-B*-expressing cells, and the boundaries of neighbouring cells, were generated using the program CorelDRAW Graphics Suite 11. Each series of images was aligned using DeltaViewer software (<http://delta.math.sci.osaka-u.ac.jp/DeltaViewer/>) to build the three-dimensional image.

18. Espina, V. *et al.* Laser-capture microdissection. *Nature Protocols* **1**, 586–603 (2006).
19. Bockman, D. E. & Cooper, M. D. Pinocytosis by epithelium associated with lymphoid follicles in the bursa of Fabricius, appendix, and Peyer's patches. An electron microscopic study. *Am. J. Anat.* **136**, 455–477 (1973).

# Identification of adult nephron progenitors capable of kidney regeneration in zebrafish

Cuong Q. Diep<sup>1,2</sup>, Dongdong Ma<sup>2,3</sup>, Rahul C. Deo<sup>1,2</sup>, Teresa M. Holm<sup>1,2</sup>, Richard W. Naylor<sup>1,2</sup>, Natasha Arora<sup>1</sup>, Rebecca A. Wingert<sup>1,2,4</sup>, Frank Bollig<sup>5</sup>, Gordana Djordjevic<sup>1</sup>, Benjamin Lichman<sup>1</sup>, Hao Zhu<sup>6</sup>, Takanori Ikenaga<sup>7</sup>, Fumihito Ono<sup>7</sup>, Christoph Englert<sup>5,8</sup>, Chad A. Cowan<sup>1,2,4,9</sup>, Neil A. Hukriede<sup>10</sup>, Robert I. Handin<sup>2,3,4</sup> & Alan J. Davidson<sup>1,2,4,†</sup>

Loss of kidney function underlies many renal diseases<sup>1</sup>. Mammals can partly repair their nephrons (the functional units of the kidney), but cannot form new ones<sup>2,3</sup>. By contrast, fish add nephrons throughout their lifespan and regenerate nephrons *de novo* after injury<sup>4,5</sup>, providing a model for understanding how mammalian renal regeneration may be therapeutically activated. Here we trace the source of new nephrons in the adult zebrafish to small cellular aggregates containing nephron progenitors. Transplantation of single aggregates comprising 10–30 cells is sufficient to engraft adults and generate multiple nephrons. Serial transplantation experiments to test self-renewal revealed that nephron progenitors are long-lived and possess significant replicative potential, consistent with stem-cell activity. Transplantation of mixed nephron progenitors tagged with either green or red fluorescent proteins yielded some mosaic nephrons, indicating that multiple nephron progenitors contribute to a single nephron. Consistent with this, live imaging of nephron formation in transparent larvae showed that nephrogenic aggregates form by the coalescence of multiple cells and then differentiate into nephrons. Taken together, these data demonstrate that the zebrafish kidney probably contains self-renewing nephron stem/progenitor cells. The identification of these cells paves the way to isolating or engineering the equivalent cells in mammals and developing novel renal regenerative therapies.

Zebrafish nephrons in the adult kidney are similar to those found in the embryonic kidney<sup>6</sup>, except that they are highly branched and drained by two central collecting ducts (Fig. 1a and Supplementary Fig. 2a–j). We confirmed that zebrafish nephron number increases with age (Fig. 1b), similar to other fish<sup>4,5</sup>. To identify the source of new nephrons in adult zebrafish, we first characterized the effects of gentamicin injection, an established nephrotoxin<sup>7</sup>. Intraperitoneal injection of gentamicin induced nephron damage, downregulated the proximal tubule marker *slc20a1a* and resulted in a failure to take up filtered 40 kDa fluorescent dextran<sup>8</sup> by 1 day post-injection (Fig. 1c–f, *n* = 6/6; Fig. 1j–k, *n* = 8/8; Supplementary Fig. 2k–p). Around 4 days post-injection, partial restoration in nephron function was observed, suggesting some nephrons recovered from the injury (Fig. 1g, l, arrow). At this stage we also detected small, but appropriately proportioned, nephrons that were dextran-positive, proliferating and basophilic, which are characteristic features of immature nephrons<sup>7</sup> (approximately 15 per kidney; Fig. 1i, l inset, n). By 15 days post-injection the damaged nephrons had recovered to near-normal levels, although immature nephrons could still be detected (Fig. 1h, m, arrow).

If the adult kidney contains nephron progenitors responsible for the formation of new nephrons, then these cells might be amenable to transplantation. To test this, we developed a transplantation assay (Fig. 2a and Supplementary Fig. 3a–k) in which recipient fish were

immunocompromised by radiation to prevent graft rejection<sup>9</sup> and then injected with gentamicin. Unpurified whole-kidney marrow cells (WKM), mostly comprising non-tubular interstitial cells<sup>9</sup>, were prepared from Tg(*cdh17:EGFP*)<sup>10</sup> or Tg(*cdh17:mCherry*) donors that express fluorescent reporters in the distal nephron. Injection of approximately  $5 \times 10^5$  of these cells resulted in donor-derived nephrons in 100% of the recipients (*n* = 6) by 18 days post-transplantation (d.p.t.), with an average of 24 donor-derived nephrons (Fig. 2b, arrow, inset). Donor nephron number increased with time, reaching an average of 70 nephrons by 59 d.p.t. (Fig. 2c) and greatly expanded the head kidney on the injected side (Fig. 2d, arrow). At these later time points, we also found donor-derived nephrons in locations distant from the site of injection, which suggests that the transplanted cells are migratory (Supplementary Fig. 3l, arrowheads).

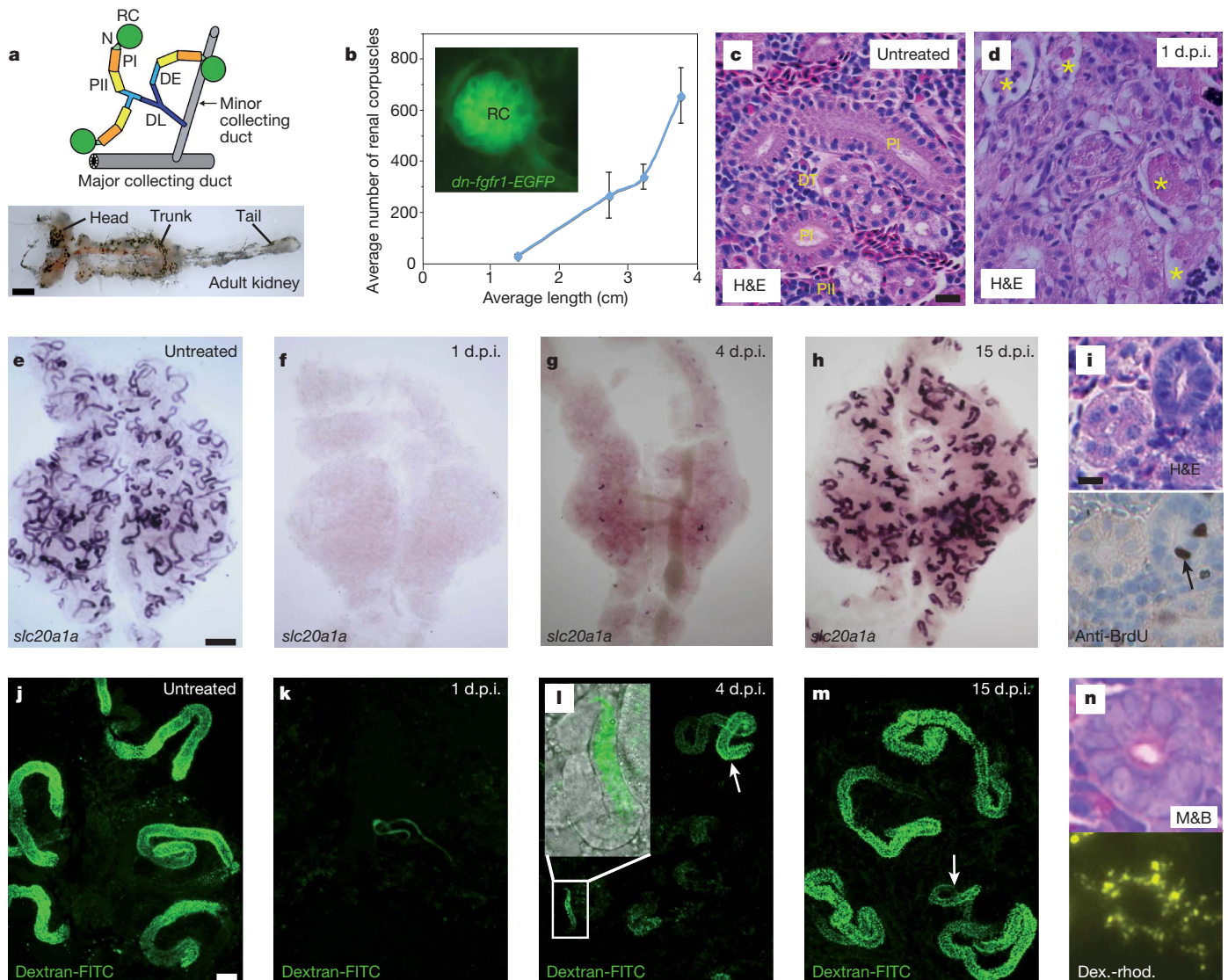
To regenerate damaged tissue successfully, newly created structures must incorporate into existing tissue. To determine whether the donor-derived nephrons were capable of blood filtration, we injected 40 kDa fluorescent dextran into transplant recipients that had received WKM from Tg(*cdh17:mCherry*) donor fish and dissected out individual nephrons. All of the donor-derived nephrons examined (*n* = 5) were dextran-positive (Fig. 2e), indicating that they had integrated into the recipient's blood supply. These results show that nephron progenitors are present in the adult kidney and that after transplantation they are capable of forming new functional nephrons within the host's renal tissue.

Cell transplantation experiments can be confounded by the fusion of donor and recipient cells. To address this, we injected WKM from Tg(*cdh17:mCherry*) donors into Tg(*cdh17:EGFP*) recipients. If fusion occurred, we would expect to find nephrons positive for both mCherry and enhanced green fluorescent protein (EGFP). An analysis of engrafted recipients (*n* = 6) revealed that all of the mCherry-positive nephrons were EGFP-negative, providing evidence that they had not formed by cell–cell fusion. In addition, we identified the connection of the donor-derived nephrons with the host's renal tubules, providing further evidence that the engrafted nephrons had successfully integrated into the recipient's renal system (Fig. 2f).

Lineage labelling studies in the developing mouse kidney have revealed that multiple *Six2*<sup>+</sup> cap mesenchyme cells, the source of nephron progenitors, contribute to a single nephron<sup>11</sup>. To explore this in zebrafish, we transplanted a 1:1 mix of Tg(*cdh17:EGFP*) and Tg(*cdh17:mCherry*) WKM cells into conditioned recipients. Mosaic nephrons containing both EGFP-positive and mCherry-positive cells were found in 27% of the engrafted fish (*n* = 15; Fig. 2g), although the remaining nephrons were either all EGFP-positive or all mCherry-positive. Thus multiple nephron progenitors can contribute to an individual nephron.

<sup>1</sup>Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>2</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Hematology Division, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>4</sup>Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. <sup>5</sup>Leibniz Institute for Age Research, Fritz Lipmann Institute, Jena D-07745, Germany. <sup>6</sup>Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>7</sup>Section on Model Synaptic Systems, Laboratory of Molecular Physiology, National Institutes of Health/National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland 20892, USA. <sup>8</sup>Friedrich-Schiller-University, Jena D-07743, Germany. <sup>9</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>10</sup>Department of Developmental Biology, University of Pittsburgh, School of Medicine, Pittsburgh, Pennsylvania 15260, USA. <sup>†</sup>Present address: Department of Molecular Medicine and Pathology, School of Medical Sciences, The University of Auckland, Auckland 1142, New Zealand.





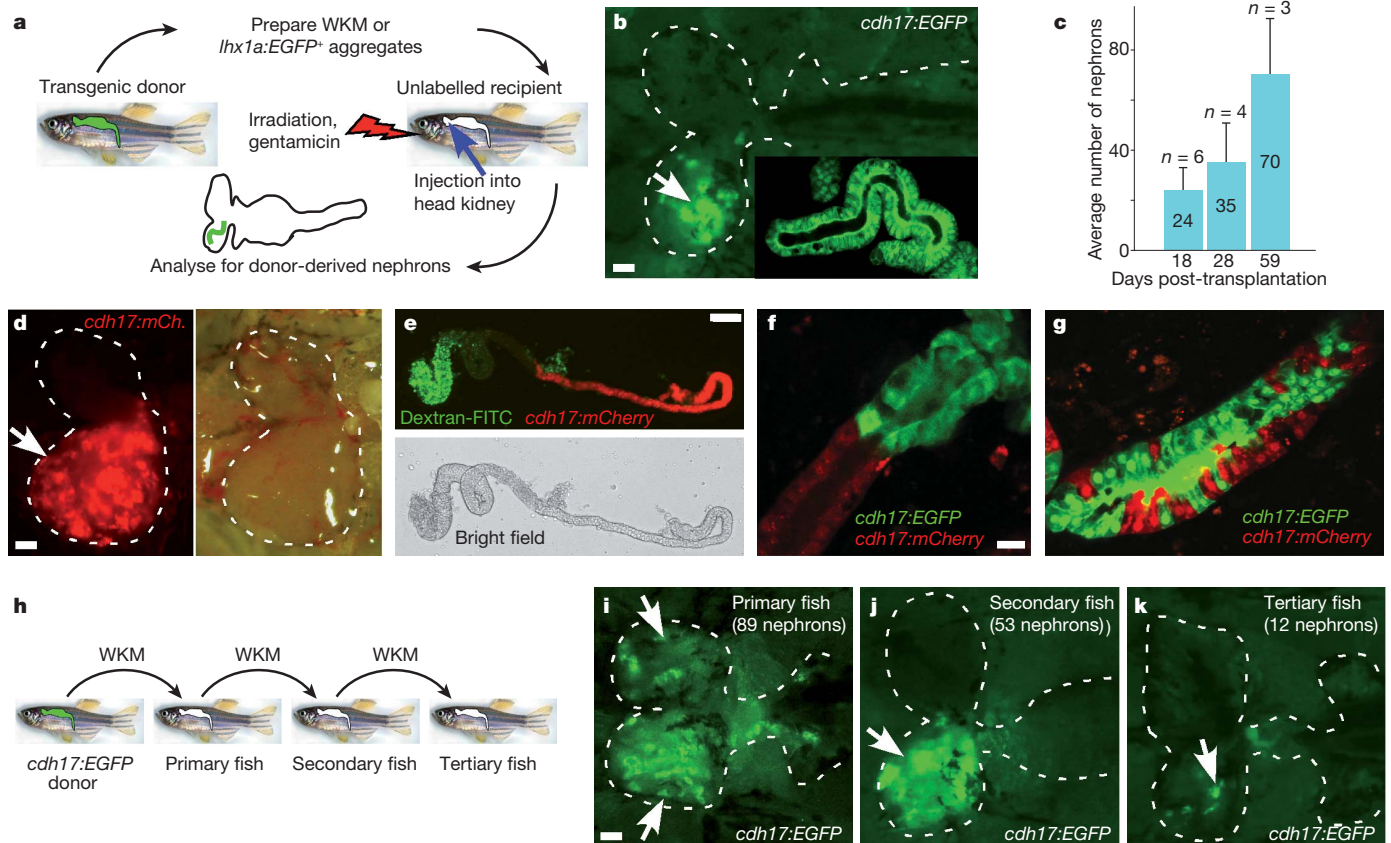
**Figure 1 | The adult zebrafish kidney undergoes nephrogenesis throughout life and after injury.** **a**, Zebrafish kidney and nephron model (scale bar, 1 mm). **b**, Graph showing average number of renal corpuscles (RC) relative to body length (inset shows an RC labelled with *dn-fgfr1-EGFP*<sup>18</sup>) (error bar, one standard deviation;  $n = 3$  fish per time point). RC, renal corpuscle; N, neck; PI, proximal tubule I; PII, proximal tubule II; DE, distal early; DL, distal late. **c, d**, Kidney sections showing gentamicin-damaged nephrons (asterisks; scale bar, 10  $\mu$ m). H&E, haematoxylin and eosin; d.p.i., days post-injection; DT,

distal tubule. **e-h**, Expression of *slc20a1a* in gentamicin-damaged kidneys (scale bar, 0.5 mm). **i**, Immature nephron with dividing cells (arrow) (scale bar, 10  $\mu$ m). **j-m**, Uptake of 40 kDa fluorescein isothiocyanate (FITC)-conjugated dextran by gentamicin-damaged kidneys (inset in **l** shows an immature nephron; arrow in **l** marks a positive nephron; arrow in **m** indicates an immature nephron; scale bar, 30  $\mu$ m). **n**, Serial sections showing that immature nephrons take up 40 kDa dextran-rhodamine. M&B, methylene blue and basic fuchsin.

Mammalian *Six2*<sup>+</sup> cap mesenchyme cells are also characterized by their stem-cell-like self-renewal properties<sup>11</sup>. Serial transplantation is used to distinguish haematopoietic stem cells from progenitors<sup>12</sup>. We investigated whether we could obtain donor-derived nephrons after serial transplantation of WKM from engrafted recipients (Fig. 2h). We transplanted WKM from primary fish containing 2–89 *cdh17:EGFP*<sup>+</sup> donor-derived nephrons and achieved a 48% ( $n = 21$ ) engraftment rate in secondary fish, with the number of donor-derived nephrons ranging from 1 to 53 by 41 d.p.t. The WKM from one of these secondary fish, containing 53 engrafted nephrons, was transplanted again and successfully engrafted a third time, giving rise to 12 donor-derived nephrons in the tertiary recipient at 35 d.p.t. (a total of 135 days from primary to tertiary fish; Fig. 2i–k). These results demonstrate that nephron progenitors possess significant proliferative potential, consistent with self-renewing capabilities.

We next sought to identify the cells responsible for nephron progenitor activity. We noted that approximately 0.1% of the WKM from

Tg(*cdh17:EGFP*) fish is EGFP-positive (Supplementary Fig. 4a). To test whether *cdh17:EGFP*<sup>+</sup> cells could contribute to new nephrons, we sorted and transplanted this fraction (approximately 5,000 cells per fish) but failed to observe engraftment ( $n = 0/7$ ). We subsequently explored other markers of nephron progenitors. In mammals, nephrogenesis initiates with the formation of ‘pre-tubular aggregates’ that undergo a mesenchymal-to-epithelial transition into renal vesicles<sup>13</sup>. These structures express several transcription factors including *Lhx1/Lim1* (ref. 14) and *Wt1* (ref. 15). We therefore examined the Tg(*lhx1a:EGFP*)<sup>16</sup> and Tg(*wt1b:mCherry*) transgenic lines to determine whether these reporters mark nephron progenitors. Kidneys from untreated Tg(*lhx1a:EGFP*) adults were found to contain three distinctive EGFP-positive cell populations: (1) single cells with a mesenchymal morphology (Fig. 3a) that make up approximately 0.02% of the WKM (Supplementary Fig. 4b), (2) homogeneous aggregates of *lhx1a:EGFP*<sup>+</sup> mesenchymal cells ranging from a few to approximately 30 cells (Fig. 3b, c and Supplementary Fig. 4i) (approximately 100 aggregates



**Figure 2 | The adult zebrafish kidney contains transplantable progenitors that form functional nephrons.** **a**, Overview of the transplantation assay. **b**, A primary transplanted fish at 18 d.p.t. with *cdh17:EGFP*<sup>+</sup> donor-derived nephrons (arrow; inset, higher magnification view; scale bar, 0.5 mm). **c**, Average number of donor-derived nephrons over time (error bar, one standard deviation; *n*, total fish per time point). **d**, Head kidney of a recipient at 34 d.p.t. showing expansion of renal tissue caused by *cdh17:mCherry*<sup>+</sup> donor-derived nephrons (arrow; scale bar, 0.5 mm). **e**, A *cdh17:mCherry*<sup>+</sup> donor-

derived nephron showing functional uptake of 40 kDa FITC-conjugated dextran (scale bar, 30  $\mu$ m). **f**, Connection of donor-derived nephrons (*cdh17:mCherry*<sup>+</sup>) with the *cdh17:EGFP*<sup>+</sup> recipient's renal system (scale bar, 10  $\mu$ m). **g**, A mosaic nephron arising from the co-injection of a mixture of *cdh17:EGFP*- and *cdh17:mCherry*-labelled nephron progenitors. **h**, Overview of the serial transplantation assay. **i-k**, Donor-derived nephrons (*cdh17:EGFP*<sup>+</sup>, arrows) in primary-, secondary- and tertiary-engrafted recipients (scale bar, 0.5 mm).

per kidney) and (3) renal vesicle-like bodies (0–2 per kidney; Fig. 3d). The last two populations were highly reminiscent of pre-tubular aggregates and renal vesicles in mammals.

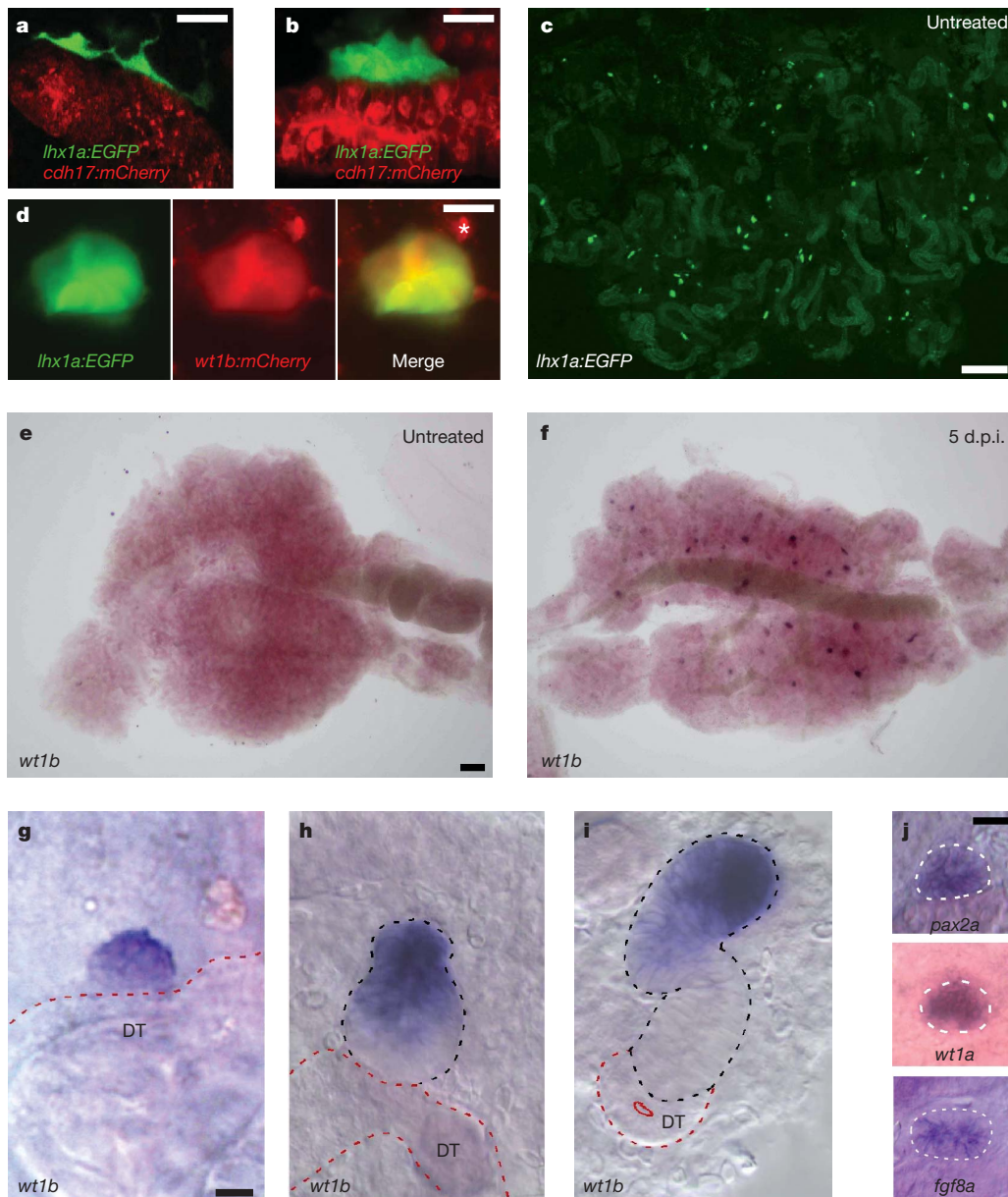
An examination of Tg(*lhx1a:EGFP*; *wt1b:mCherry*) double transgenic kidneys revealed that only the large aggregates and renal vesicles, but not the other *lhx1a:EGFP*<sup>+</sup> populations, were *wt1b:mCherry*<sup>+</sup> (Fig. 3d). We hypothesized that the *lhx1a:EGFP*<sup>+</sup>/*wt1b:mCherry*<sup>+</sup> renal vesicle-like bodies, which were rare in untreated kidneys, constitute primitive nephrons. Consistent with this, gentamicin treatment greatly induced the formation of *lhx1a:EGFP*<sup>+</sup>/*wt1b:mCherry*<sup>+</sup> double-positive cells (data not shown) and activated the endogenous expression of the early-acting renal genes *pax2a*, *fgf8a*, *wt1a* and *wt1b* in similar structures (Fig. 3e–j and Supplementary Fig. 4c–h). We failed to detect the expression of mature nephron markers in structures resembling either *lhx1a:EGFP*<sup>+</sup> aggregates or *wt1b*<sup>+</sup> renal vesicles (Supplementary Fig. 5a–c and Supplementary Table 1). Similarly, quantitative PCR analysis of purified *lhx1a:EGFP*<sup>+</sup> and *cdh17:EGFP*<sup>+</sup> cells showed that *lhx1a:EGFP*<sup>+</sup> cells express considerably lower levels of mature nephron markers than *cdh17:EGFP*<sup>+</sup> cells (Supplementary Fig. 5d). These findings suggest that *lhx1a:EGFP* labels nephron progenitors and *lhx1a:EGFP*/*wt1b:mCherry* labels early-stage nephrons.

To clarify the lineage relationships between *lhx1a:EGFP*<sup>+</sup> and *lhx1a:EGFP*<sup>+</sup>/*wt1b:mCherry*<sup>+</sup> cells, we took advantage of the optical transparency of larval fish to visualize nephrogenesis *in vivo*. By observing Tg(*cdh17:EGFP*) larvae as well as using wholemount *in situ* hybridization, we found that adult kidney formation initiates at the 5.2-mm stage (approximately 13 days post-fertilization). The first

nephron appears consistently on the embryonic (pronephric) renal tubules just posterior to the swim bladder (Supplementary Fig. 6a–e and data not shown). *lhx1a:EGFP*<sup>+</sup> cells appeared before this, at the 4-mm stage (approximately 10 days post-fertilization) (Fig. 4a, arrow, inset), rapidly migrated along the pronephric tubules (Fig. 4b, arrows), and formed into aggregates (Fig. 4b, arrowheads). An *in vivo* time course of Tg(*lhx1a:EGFP*; *cdh17:mCherry*) and Tg(*lhx1a:EGFP*; *wt1b:mCherry*) larvae showed that the *lhx1a:EGFP*<sup>+</sup> aggregates arose from the coalescence of three or four *lhx1a:EGFP*<sup>+</sup> cells that expanded to form a renal vesicle and activated expression of *wt1b:mCherry* (Fig. 4d and Supplementary Fig. 6f). Similar time courses of Tg(*lhx1a:EGFP*; *cdh17:mCherry*) and Tg(*wt1b:EGFP*; *pax8:DsRed*) larvae demonstrated that the renal vesicle elongated into a *cdh17*<sup>+</sup> nephron, with *lhx1a*<sup>+</sup> cells becoming restricted to the point of fusion with the pronephric tubules, *pax8* initiating in the distal tubule and *wt1b* labelling the glomerulus and proximal tubule (Fig. 4e and Supplementary Fig. 6g). To demonstrate a requirement of *lhx1a:EGFP*<sup>+</sup> cells for nephrogenesis, we ablated single *lhx1a:EGFP*<sup>+</sup> aggregates with a laser (Fig. 4c, arrow), resulting in aborted nephrogenesis in the targeted region without affecting neighbouring nephrons (Fig. 4c, arrowhead) (*n* = 2/2).

Next we tested whether *lhx1a:EGFP*<sup>+</sup> cells had nephron-forming activity. Transplantation of single *lhx1a:EGFP*<sup>+</sup> cells failed to engraft conditioned recipients. However, transplantation of individual *lhx1a:EGFP*<sup>+</sup> aggregates resulted in successful engraftment in 33% (*n* = 15) of transplanted fish (Fig. 4f, g). In one case, a single aggregate contributed to 16 nephrons, 27 aggregates and numerous individual cells (Supplementary Fig. 7a–c), consistent with *lhx1a:EGFP*<sup>+</sup> cells





**Figure 3 | Expression of *lhx1a:EGFP* and other renal factors in the adult kidney.** **a**, Single mesenchymal cells labelled with *lhx1a:EGFP* (scale bar, 10  $\mu$ m). **b**, Small aggregates labelled with *lhx1a:EGFP* (scale bar, 10  $\mu$ m). **c**, An untreated kidney showing *lhx1a:EGFP*<sup>+</sup> aggregates in a portion of the trunk region; scale bar, 0.5 mm). **d**, Co-expression of *wt1b:mCherry* and *lhx1a:EGFP* in large aggregates (asterisk, autofluorescence; scale bar, 10  $\mu$ m). **e**, **f**, Expression

having extensive proliferative and self-renewing capabilities. Transplantation of *lhx1a:EGFP*<sup>+</sup>/*wt1b:mCherry*<sup>+</sup> renal vesicles failed to engraft ( $n = 0/10$ ), suggesting that nephron-forming potential is restricted to *lhx1a:EGFP*<sup>+</sup> aggregates. These findings demonstrate that *lhx1a:EGFP*<sup>+</sup> aggregates contain nephron progenitors and support our observation that multiple nephron progenitors are needed to form a nephron.

To determine how similar *lhx1a:EGFP*<sup>+</sup> cells are to *Six2*<sup>+</sup> mouse cap mesenchyme cells, we conducted a microarray analysis, comparing the genes upregulated in *lhx1a:EGFP*<sup>+</sup> cells (relative to *cdh17:EGFP*<sup>+</sup> epithelial cells) with those upregulated in mouse *Six2*<sup>+</sup> cells (relative to mouse proximal tubule epithelial cells). At a global level, the respective gene sets that are upregulated in *lhx1a:EGFP*<sup>+</sup> cells and *Six2*<sup>+</sup> cells are not significantly similar (Supplementary Tables 2–4 and Supplementary Fig. 8a). However, there is conservation of several factors implicated in renal development and/or stem-cell self-renewal. Notably,

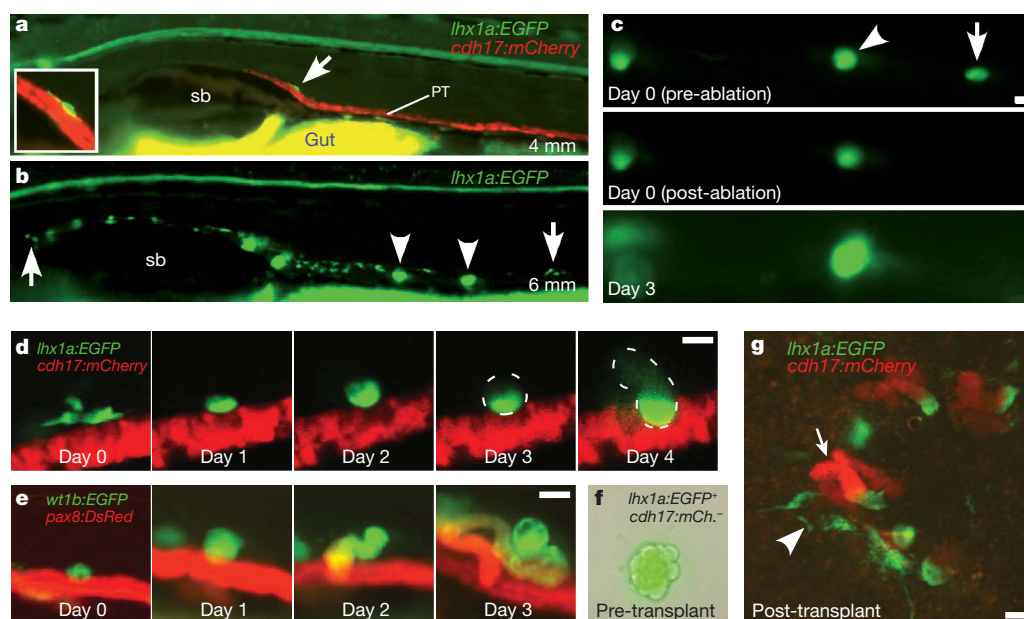
of *wt1b* in untreated and gentamicin-damaged kidneys (scale bar, 0.5 mm).

**g–i**, Expression of *wt1b* in a large aggregate (**g**), a comma-shaped body (**h**) and an immature nephron (**i**) in gentamicin-damaged kidneys (scale bar, 10  $\mu$ m). **j**, Expression of *pax2a*, *wt1a* and *fgf8a* in large aggregates or renal vesicles in gentamicin-damaged kidneys (scale bar, 10  $\mu$ m).

orthologues of *Six2* (*six2a*) and *Wt1* (*wt1a*), which are essential for cap mesenchyme maintenance, are upregulated both in *lhx1a:EGFP*<sup>+</sup> cells and in *Six2*<sup>+</sup> cells. Quantitative PCR confirmed that *six2a* and *wt1a* are expressed over 15-fold higher in *lhx1a:EGFP*<sup>+</sup> cells than *cdh17:EGFP*<sup>+</sup> cells (Supplementary Fig. 8b). Several other potentially important regulators were also identified in the comparison, including *Meis2*, *Ezh2* and *Tcf3*, which are implicated in Wnt signalling and/or stem-cell function (Supplementary Table 4). These results suggest that, despite having distinct molecular identities, zebrafish *lhx1a:EGFP*<sup>+</sup> cells and *Six2*<sup>+</sup> cells share a core set of regulatory genes that may be important for conferring renal stem/progenitor cell potential.

In conclusion, we have identified an adult population of nephron progenitors that reside in small aggregates throughout the zebrafish kidney. These cells are uniquely defined by their ability to form new functional nephrons during zebrafish growth, injury and after transplantation. Nephron progenitors can be serially transplanted, consistent





**Figure 4** | *lhx1a:EGFP*<sup>+</sup> cells form nephrons during adult kidney development and after transplantation. **a**, Lateral view of a Tg(*lhx1a:EGFP;cdh17:mCherry*) larva showing the first *lhx1a:EGFP*<sup>+</sup> cell to appear on top of the *cdh17:mCherry*<sup>+</sup> embryonic kidney tubules (arrow and inset). **b**, Lateral view of a Tg(*lhx1a:EGFP*) larva showing the extent of *lhx1a:EGFP*<sup>+</sup> cell migration (arrows) and their aggregation (arrowheads). **c**, Laser-ablation of an *lhx1a:EGFP*<sup>+</sup> aggregate (arrow) inhibits nephron formation without affecting nephrogenesis of an adjacent aggregate (arrowhead). **d**, Time course of a Tg(*lhx1a:EGFP;cdh17:mCherry*) larva

demonstrating that *lhx1a:EGFP*<sup>+</sup> cells coalesce into an aggregate and differentiate into a nephron. **e**, Time course of a Tg(*wt1b:EGFP;pax8:DsRed*) larva showing development of a *wt1b:EGFP*<sup>+</sup> aggregate into a nephron. **f**, Bright field and fluorescent merge of an aggregate from a Tg(*lhx1a:EGFP;cdh17:mCherry*) donor. **g**, Donor-derived *cdh17:mCherry*<sup>+</sup> nephrons (one indicated by arrow) and multiple single *lhx1a:EGFP*<sup>+</sup> cells (arrowhead) resulting from the transplantation of the aggregate shown in **f** (scale bar, 30  $\mu$ m). PT, pronephric tubule; sb, swim bladder; larvae shown with anterior to the left.

with stem-cell capabilities, although confirmation of this awaits direct lineage-tracing experiments. Our *in vivo* imaging of nephrogenesis and chimaeric transplantation results demonstrated that nephrogenic aggregates form by the coalescence of multiple *lhx1a:EGFP*<sup>+</sup> cells (Supplementary Fig. 1). This process is reminiscent of nephrogenesis in mammals and suggests that similar mechanisms govern nephron formation in both species. Consistent with this, *lhx1a:EGFP*<sup>+</sup> cells express *six2a* and *wt1a*, two critical regulators of mammalian nephron progenitors. Our observation that only aggregates of *lhx1a:EGFP*<sup>+</sup> cells, but not single cells, are capable of engraftment suggests that nephron progenitor potential may depend upon a 'community effect'<sup>17</sup>, a phenomenon whereby continued cell contact is necessary for cells to respond to an inductive signal. The failure of renal vesicles to engraft suggests that nephron-forming potential is lost upon epithelial differentiation.

With our data in hand, it is now possible to pursue whether the mammalian adult kidney contains an equivalent population of nephrogenic aggregates. If present, these cells are most probably dormant or their regenerative abilities blocked, given that nephrogenesis ceases around birth. Using zebrafish to understand the molecular identity of nephron progenitors and the pathways that regulate them may lead to therapeutic ways to activate, or artificially engineer, the mammalian counterpart and augment human renal regeneration. With the rise in chronic kidney disease becoming a serious worldwide healthcare issue, a nephron-progenitor-based regenerative therapy will have a major clinical impact.

## METHODS SUMMARY

For WKM transplants, adult recipient fish were conditioned with intraperitoneal injection of gentamicin (80  $\mu$ g g<sup>-1</sup>), then immunocompromised with sub-lethal  $\gamma$ -irradiation (25 Gy) to prevent graft rejection<sup>9</sup>. Unpurified WKM cells were prepared as previously described<sup>9</sup> from Tg(*cdh17:EGFP*)<sup>10</sup> or Tg(*cdh17:mCherry*) donors that express fluorescent reporter proteins in the distal nephron. For the *lhx1a:EGFP*<sup>+</sup> single cell and aggregate transplants, dissected kidneys from

Tg(*lhx1a:EGFP;cdh17:mCherry*) fish were treated with 10% collagenase/dispase for 15 min and cells/aggregates manually transferred with a mouth pipette to a drop of 1X PBS/2% fetal calf serum on a glass slide. A single cell or aggregate was serially passaged through three droplets of PBS/fetal calf serum until free of non-positive cells just before transplantation.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 April; accepted 15 November 2010.

Published online 26 January 2011.

1. Fogo, A. B. Mechanisms of progression of chronic kidney disease. *Pediatr. Nephrol.* **22**, 2011–2022 (2007).
2. Humphreys, B. D. *et al.* Intrinsic epithelial cells repair the kidney after injury. *Cell Stem Cell* **2**, 284–291 (2008).
3. Hartman, H. A., Lai, H. L. & Patterson, L. T. Cessation of renal morphogenesis in mice. *Dev. Biol.* **310**, 379–387 (2007).
4. Reimschuessel, R. A fish model of renal regeneration and development. *ILAR J.* **42**, 285–291 (2001).
5. Zhou, W. *et al.* Characterization of mesonephric development and regeneration using transgenic zebrafish. *Am. J. Physiol. Renal Physiol.* **299**, F1040–F1047 (2010).
6. Wingert, R. A. *et al.* The *cdx* genes and retinoic acid control the positioning and segmentation of the zebrafish pronephros. *PLoS Genet.* **3**, 1922–1938 (2007).
7. Augusto, J., Smith, B., Smith, S., Robertson, J. & Reimschuessel, R. Gentamicin-induced nephrotoxicity and nephroneogenesis in *Oreochromis nilotica*, a tilapia fish. *Dis. Aquat. Organ.* **26**, 49–58 (1996).
8. Kramer-Zucker, A. G., Wiessner, S., Jensen, A. M. & Drummond, I. A. Organization of the pronephric filtration apparatus in zebrafish requires Nephron, Podocin and the FERM domain protein Mosaic eyes. *Dev. Biol.* **285**, 316–329 (2005).
9. Traver, D. *et al.* Effects of lethal irradiation in zebrafish and rescue by hematopoietic cell transplantation. *Blood* **104**, 1298–1305 (2004).
10. Hall, C. *et al.* Transgenic zebrafish reporter lines reveal conserved Toll-like receptor signaling potential in embryonic myeloid leukocytes and adult immune cell lineages. *J. Leukoc. Biol.* **85**, 751–765 (2009).
11. Kobayashi, A. *et al.* Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development. *Cell Stem Cell* **3**, 169–181 (2008).
12. Purton, L. E. & Scadden, D. T. Limiting factors in murine hematopoietic stem cell assays. *Cell Stem Cell* **1**, 263–270 (2007).
13. Dressler, G. R. The cellular basis of kidney development. *Annu. Rev. Cell Dev. Biol.* **22**, 509–529 (2006).

14. Kobayashi, A. *et al.* Distinct and sequential tissue-specific activities of the LIM-class homeobox gene *Lim1* for tubular morphogenesis during kidney development. *Development* **132**, 2809–2823 (2005).
15. Georgas, K. *et al.* Analysis of early nephron patterning reveals a role for distal RV proliferation in fusion to the ureteric tip via a cap mesenchyme-derived connecting segment. *Dev. Biol.* **332**, 273–286 (2009).
16. Swanhart, L. *et al.* Characterization of an *lhx1a* transgenic reporter in zebrafish. *Int. J. Dev. Biol.* **54**, 731–736 (2010).
17. Gurdon, J. B. A community effect in animal development. *Nature* **336**, 772–774 (1988).
18. Lee, Y., Grill, S., Sanchez, A., Murphy-Ryan, M. & Poss, K. D. Fgf signaling instructs position-dependent growth rate during zebrafish fin regeneration. *Development* **132**, 5173–5183 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. C. Liao for help with suturing, and R. Ethier and L. Gyr for zebrafish care. A.J.D. was supported by the Harvard Stem Cell Institute, the

American Society of Nephrology and the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases (P50DK074030).

**Author Contributions** C.Q.D. and A.J.D. designed the experimental strategy, analysed data, prepared the manuscript, and generated and characterized the *Tg(cdh17:EGFP)*, *Tg(cdh17:mCherry)* and *Tg(wt1b:mCherry)* lines. C.Q.D. performed the regeneration, transplants, time course and ablation experiments. C.Q.D., D.M. and R.I.H. made the initial observation that nephron progenitors can be transplanted. N.A.H. generated the *Tg(lhx1a:EGFP)* line (R01DK069403), F.B. and C.E. generated the *Tg(wt1b:EGFP)* line, and T.I. and F.O. provided the *Tg(pax8:DsRed)* line. N.A., R.A.W., G.D. and B.L. analysed kidney expression. H.Z. provided sections of regenerating kidneys. R.C.D., T.M.H., R.W.N., and C.A.C. performed quantitative PCR and microarray analyses. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.J.D. ([a.davidson@auckland.ac.nz](mailto:a.davidson@auckland.ac.nz)).

## METHODS

**Zebrafish transgenic lines.** Zebrafish were maintained as previously described<sup>19</sup> and according to Institutional Animal Care and Use Committee protocols. The transgenic lines Tg(*wt1b:EGFP*), Tg(*lhx1a:EGFP*) and Tg(*hsp70:dn-fgfr1-EGFP*) were previously reported<sup>16,18,20</sup>. The Tg(*wt1b:mCherry*) line was generated by replacement of EGFP with mCherry in the F47 vector, which contains a shortened version of the *wt1b* promoter that was previously described<sup>20</sup>. The Tg(*cdh17:EGFP*) line was generated by isolation of a 4.3-kb genomic fragment upstream of the Exon 1 5' untranslated region. The promoter fragment was cloned into the XhoI/SalI sites of the Tol2 vector T2AL200R150G<sup>21</sup>. The Tg(*cdh17:mCherry*) line was generated by co-injection of Cre mRNA with the Tol2 vector T2cdh17-loxP-EGFP-loxP-mCherry. The Tg(*pax8:DsRed*) line was generated by gene trap screening. DsRedExpress was inserted into the BamHI/NotI sites of the pT2KSAG Tol2 vector<sup>22</sup> and was used to generate the transgenic line. Mapping of the insertion site by inverse PCR revealed that DsRedExpress was inserted in the intron region between exons 1 and 2 (T.I. and F.O., unpublished observations).

**Adult and larval zebrafish experiments.** Epifluorescent images were taken from a Nikon Eclipse 80i microscope using the Hamamatsu ORCA-ER camera and confocal images were acquired using the A1 high-speed confocal Ti-e inverted microscope system (Nikon).

Adult. Gentamicin (40 µg), BrdU (100 µg) and 40 kDa dextran-FITC or -rhodamine (2 µg) were administered by intraperitoneal injection. Single nephrons were dissected in Ringer's buffer using sharpened tungsten needles. Kidney wholemount *in situ* hybridization was as previously described (<http://zfin.org/ZFIN/Methods/ThisseProtocol.html>) with the addition of 1% dimethyl sulphoxide supplemented to the fixative. Fluorescence-activated cell sorting and analyses were performed using the BD FACSAria (Harvard Stem Cell Institute Flow Cytometry Core Facility). For transplantation of cells directly into the kidney, conditioned recipients were anaesthetized with 0.02% tricaine and a lateral incision was made posterior to the gills and level with the kidney. One microlitre containing approximately  $5 \times 10^5$  WKM cells, or a single *lhx1a:EGFP*<sup>+</sup> cell or a single *lhx1a:EGFP*<sup>+</sup> aggregate was injected directly into the head kidney of recipient fish (3 days after conditioning) using a Hamilton syringe. The incision was closed with a suture and the fish was returned to water (Supplementary Fig. 3a–k).

Larva. Larval wholemount *in situ* hybridization was performed as reported<sup>23</sup>. Ablation of *lhx1a:EGFP*<sup>+</sup> aggregates was performed with the MicroPoint Laser System (Photonic Instruments) in conjunction with the Nikon Eclipse 80i microscope. For the cellular necrosis assay, water control and gentamicin-treated kidneys were incubated for 10 min in  $5 \mu\text{g ml}^{-1}$  of acridine orange (Sigma) in PBS, washed three times with 50 ml PBS and imaged under bright field and epifluorescence (FITC).

**Histology.** Haematoxylin and eosin. Kidneys and larvae were fixed in 4% paraformaldehyde, embedded in paraffin, sectioned and stained with haematoxylin and eosin or antibodies against EGFP and BrdU (Dana-Farber/Harvard Cancer Center Pathology Core Facility).

Methylene blue and basic fuchsin. Kidneys were fixed in 4% paraformaldehyde, embedded in JB4 resin, sectioned and stained with methylene blue and basic fuchsin. **Microarray analysis.** Triplicate samples of approximately 4,000 *lhx1a:EGFP*<sup>+</sup> and *cdh17:EGFP*<sup>+</sup> cells were sorted by fluorescence-activated cell sorting into lysis buffer, complementary DNAs were amplified, labelled with Cy3 (*cdh17:EGFP*<sup>+</sup>) and Cy5 (*lhx1a:EGFP*<sup>+</sup>), and hybridized to the Agilent Whole Zebrafish Genome Oligo Microarray (3 × 44k) by Miltenyi Biotec. All statistical analysis used the R package 2.9.2 (<http://cran.r-project.org/>). Agilent microarrays were processed using the Agilent Feature Extraction software to obtain intensity ratios for each of the 43,803 probes on the array. Intensity ratios from the three separate Agilent

microarrays were subsequently quantile normalized using the normalize Between Arrays function in the affy package and differentially expressed gene between cases and controls determined using the limma package. Given the multiple number of hypotheses tested we selected the qvalue package for false discovery rate (FDR) estimation. RMA normalization, limma and qvalue were used to identify differentially expressed probe sets across the eight GUDMAP mouse kidney microarrays of interest (three for *Six2*<sup>+</sup> cells, five for proximal tubule cells), which were downloaded from the GEO database (GSE12588, GSE6589 and GSE6290). To compare the differentially expressed genes from the two platforms (and species), the microarray probes were matched to gene identities using annotation files provided by the manufacturers. Zebrafish gene identities were then mapped to mouse orthologues using a combination of the InParanoid (<http://inparanoid.sbc.su.se/>) and Ensembl databases. A total of 11,870 Ensembl mouse protein identities were identified that could be compared from both arrays. Given the smaller number of zebrafish arrays, FDR thresholds of 40% (nominal  $P < 0.016$ ), 45% ( $P < 0.048$ ) and 50% ( $P < 0.11$ ) were combined with a minimum twofold change in intensity, which corresponded to 1,635, 3,335 and 5,879 probes, respectively. More stringent FDR thresholds of 5% (nominal  $P < 0.022$ ), 10% ( $P < 0.07$ ) and 15% ( $P < 0.14$ ) were chosen for the mouse probes, leading to 5,250, 5,699, 5,936 probe sets, respectively. Genes that were both significantly upregulated and downregulated in the analysis, as judged by separate probes, were excluded, leaving a total of 10,421 genes for comparison. We used Fisher's exact test for statistical significance over a range of FDR values, yielding a range from  $P = 0.051$  to  $P = 0.86$ . The GEO database record number for the zebrafish microarray data is GSE24803.

**Quantitative PCR.** *lhx1a:EGFP*<sup>+</sup> and *cdh17:EGFP*<sup>+</sup> cells (300–9,000) were purified by fluorescence-activated cell sorting, lysed and complementary DNA generated using the Cells-to-cDNA kit (Ambion). Quantitative PCR was performed in triplicate using SYBR Green chemistry on a Mastercycler RealPlex<sup>2</sup> PCR machine (Eppendorf) using the following primer sets: *slc20a1a* forward 5'-TCTCTGGACACATTGCATC-3', reverse 5'-AGCAGTTCAGCCATTGAC-3'; *slc13a1* forward 5'-TGCTGGGATTCTGTTCTTC-3', reverse 5'-AAACCTCCACCAACAAGCAG-3'; *slc12a1* forward 5'-TCAACGCTCTGAAGAAGCTG-3', reverse 5'-ACGTTGTGTGGGTTTCTTC-3'; *slc12a3* forward 5'-ACAGATCCGGCTGAATGAAG-3', reverse 5'-AGCCAAGCCATGTAAAGAGG-3'; *six2a* forward 5'-AGCTCGGAGGATGAGTTTTC-3', reverse 5'-ATGGTGCCTTGCAGAAGAG-3'; *wt1a* forward 5'-AGCCAACCAAGGATGTTCAG-3', reverse 5'-AACC TTGATTCCTGGAGCTG-3'. *ef1a* was used as the normalization control: forward 5'-CTGGAGGCCAGCTCAAACAT-3', reverse 5'-ATCAAGAAGAGTAGTACCGCTAGCATTAC-3'. Relative quantification of target gene expression was evaluated using the comparative  $C_T$  method. A mean and standard deviation were determined for the  $\Delta C_T$  value for all genes of interest. The error in fold-change was obtained by considering the effect of an increase or decrease of one standard deviation in  $\Delta C_T$  value.

19. Westerfield, M. *The Zebrafish Book: A Guide for the Laboratory Use of Zebrafish* (Danio rerio) 4th edn (Univ. Oregon Press, 2000).
20. Perner, B., Englert, C. & Bollig, F. The Wilms tumor genes *wt1a* and *wt1b* control different steps during formation of the zebrafish pronephros. *Dev. Biol.* **309**, 87–96 (2007).
21. Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649 (2006).
22. Kawakami, K. *et al.* A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. *Dev. Cell* **7**, 133–144 (2004).
23. Elizondo, M. R. *et al.* Defective skeletogenesis with kidney stone formation in dwarf zebrafish mutant for *trpm7*. *Curr. Biol.* **15**, 667–671 (2005).



# Reversing pathological neural activity using targeted plasticity

Navzer D. Engineer<sup>1,2</sup>, Jonathan R. Riley<sup>1</sup>, Jonathan D. Seale<sup>1</sup>, Will A. Vrana<sup>1</sup>, Jai A. Shetake<sup>1</sup>, Sindhu P. Sudanagunta<sup>1</sup>, Michael S. Borland<sup>1</sup> & Michael P. Kilgard<sup>1</sup>

**Brain changes in response to nerve damage or cochlear trauma can generate pathological neural activity that is believed to be responsible for many types of chronic pain and tinnitus<sup>1–3</sup>. Several studies have reported that the severity of chronic pain and tinnitus is correlated with the degree of map reorganization in somatosensory and auditory cortex, respectively<sup>1,4</sup>. Direct electrical or transcranial magnetic stimulation of sensory cortex can temporarily disrupt these phantom sensations<sup>5</sup>. However, there is as yet no direct evidence for a causal role of plasticity in the generation of pain or tinnitus. Here we report evidence that reversing the brain changes responsible can eliminate the perceptual impairment in an animal model of noise-induced tinnitus. Exposure to intense noise degrades the frequency tuning of auditory cortex neurons and increases cortical synchronization. Repeatedly pairing tones with brief pulses of vagus nerve stimulation completely eliminated the physiological and behavioural correlates of tinnitus in noise-exposed rats. These improvements persisted for weeks after the end of therapy. This method for restoring neural activity to normal may be applicable to a variety of neurological disorders.**

Damage to the peripheral nervous system causes plasticity in multiple regions of the central nervous system. Significant changes have been reported in map organization, spontaneous activity, neural synchronization and stimulus selectivity<sup>2</sup>. The ideal method of testing whether map plasticity or some other form of plasticity is directly responsible for chronic pain and tinnitus would be to reverse the plasticity and evaluate the perceptual consequence.

Recent attempts to use sensory exposure or discrimination training to reverse the map changes in individuals with tinnitus or chronic pain have provided some temporary relief<sup>6,7</sup>. Although the clinical benefits were limited, these studies provide some support for the hypothesis that neural plasticity could be used to treat these conditions. It is possible that a long-lasting reversal of the pathological plasticity in these patients would provide significant relief.

Studies in animals have shown that repeatedly pairing sensory stimuli with electrical stimulation of the cholinergic nucleus basalis generates powerful and long-lasting changes in cortical organization<sup>8</sup>. In principle, this method could be used to reverse the effect of pathological plastic changes that are associated with tinnitus and chronic pain<sup>1–3,6</sup>. However, nucleus basalis stimulation is highly invasive and, thus, not practical for clinical use. We have developed a less invasive method for generating targeted neural plasticity by pairing vagus nerve stimulation (VNS) with sensory inputs, and have demonstrated a potential clinical application.

VNS triggers the release of neuromodulators known to promote plastic changes. The efficacy of VNS in enhancing plasticity seems to lie in the synergistic action of multiple neuromodulators acting in the cerebral cortex and other brain regions<sup>9</sup>. VNS improves learning and memory of associated events in rats and humans using identical VNS parameters<sup>10</sup>.

Our study tests the hypothesis that the pairing of VNS with tones could be used to drive neural plasticity that would reverse the behavioural correlate of tinnitus in noise-exposed rats. The first set of experiments confirms that repeatedly pairing a single tone frequency with VNS is sufficient to generate specific and long-lasting changes in cortical maps.

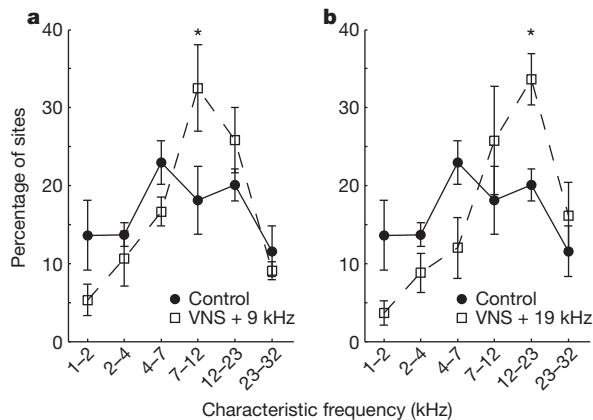
The rationale for our tinnitus therapy is that increasing the number of cortical neurons tuned to frequencies other than the tinnitus frequency ought to reduce the overrepresented tinnitus frequency. The second set of experiments confirms that repeatedly pairing a range of tone frequencies with VNS can be used to reverse the behavioural and neural correlates of tinnitus in noise-exposed rats.

In our first set of experiments, we sought to evaluate whether pairing VNS with tones can generate precise, long-lasting and large-scale changes in the frequency representation in the cortex, as we found for nucleus basalis stimulation. We paired VNS with a 9-kHz, 60-dB SPL tone ( $n = 8$  rats) or a 19-kHz, 50-dB SPL tone ( $n = 5$  rats) for 20 days (SPL, sound pressure level), 300 times per day in normal-hearing rats with cuff electrodes implanted on the left cervical vagus nerve (Methods). The VNS–tone pairing procedure was identical to earlier tone pairing procedures with nucleus basalis, ventral tegmentum or locus coeruleus stimulation that generate long-lasting map plasticity<sup>8,11,12</sup>. VNS parameters (30 Hz, 0.8 mA) were similar to the parameters used in previous rat and human VNS studies, except that the duration of stimulation and the widths of individual pulses were reduced by 60-fold and fivefold, respectively (Methods and Supplementary Fig. 2). The 0.5 s of VNS used in this study was sufficient to reduce the amplitude of the cortical electroencephalogram briefly (Supplementary Fig. 3 and supplemental data). Twenty-four hours after the last VNS–tone pairing session, we used standard micro-electrode mapping techniques to document frequency map plasticity. VNS–tone pairing caused a 70–79% increase in the number of primary auditory cortex (A1) sites with a characteristic frequency near the paired tone frequency (Fig. 1). This result confirms our hypothesis that VNS–tone pairing can be used to direct map plasticity lasting more than 24 h.

Pairing VNS with sensory stimuli is a potentially attractive method of modifying neural circuits without significant side effects. VNS is well tolerated in the 50,000 patients who currently receive VNS therapy for epilepsy or depression<sup>13</sup>. By pairing tones with brief trains of VNS, we have been able to alter cortical frequency maps significantly in rats using only 1% of the VNS that is delivered clinically (that is, 30 s every 5 min, 24 h per day) for epilepsy treatment in humans.

Having demonstrated that VNS can be used to generate specific and long-lasting map plasticity, in our second set of studies we sought to evaluate whether VNS-directed plasticity could be adapted to renormalize pathological plasticity and eliminate tinnitus. Exposure to intense, high-frequency noise is known to generate an overrepresentation of mid-frequency tones, degrade frequency selectivity and increase excitability and synchronization of auditory neurons<sup>14–16</sup>. We induced noise trauma by exposing rats to 1 h of 115-dB SPL, octave-band noise centred at 16 kHz (ref. 17; Methods). Auditory brainstem responses were used to confirm the effects of the noise exposure on hearing threshold, including temporary deafness for frequencies above 8 kHz and a long-lasting increase of auditory brainstem response thresholds and latency<sup>18</sup> (Supplementary Figs 4 and 5). After noise exposure, twice as many A1 recording sites were tuned to frequencies between 2 and 4 kHz in comparison with naive controls ( $35 \pm 7\%$  versus  $14 \pm 2\%$ ,  $P < 0.05$ ), and very few neurons responded to frequencies above

<sup>1</sup>Cortical Plasticity Laboratory, Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, Texas 75080, USA. <sup>2</sup>MicroTransponder Inc., 2802 Flintrock Trace, Suite 225, Austin, Texas 78738, USA.

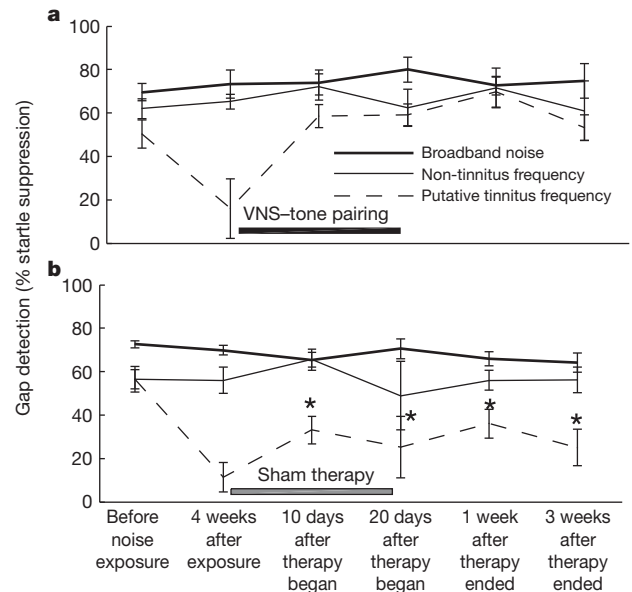


**Figure 1 | VNS-tone pairing causes map plasticity.** Repeatedly pairing VNS with a tone increases the number of A1 recordings sites tuned to the paired frequency. **a**, VNS was paired with a 9-kHz tone 6,000 times over 20 days in eight rats. **b**, VNS was paired with a 19-kHz tone in five rats. This group heard 4-kHz tones equally often but without VNS pairing. Asterisks indicate significant ( $P < 0.05$ ) increases in the fraction of A1 sites with characteristic frequencies near the paired tone. Error bars, s.e.m. This result in normal-hearing rats suggested that VNS-tone pairing might be used to reverse the map distortions induced by exposure to intense noise.

23 kHz ( $1.7 \pm 1\%$  versus  $11.5 \pm 3\%$ ,  $P < 0.01$ ). The average frequency bandwidth of A1 neurons increased by 21% ( $1.75 \pm 0.04$  versus  $1.47 \pm 0.03$  octaves at 10 dB above threshold,  $P < 0.00001$ ), and the average number of spikes evoked by a tone within each site's receptive field increased by 30% ( $4.3 \pm 0.1$  versus  $3.3 \pm 0.1$ ,  $P < 0.00001$ ). The average spontaneous rate increased by 23% ( $17.7 \pm 0.6$  versus  $14.3 \pm 0.4$  Hz,  $P < 0.00001$ ). The degree of synchronization during silence measured using the correlation coefficient between multiunit activity recorded at nearby sites was significantly increased ( $1.7 \pm 0.01$  versus  $0.19 \pm 0.01$  synchronous spikes per second of silence,  $P < 0.05$ ; Methods). These changes in frequency tuning and synchronization are similar to the physiological changes observed after noise exposure that have been proposed to be directly responsible for tinnitus<sup>2,19</sup>. Earlier studies using several different methods have documented that noise exposure can generate behavioural correlates of tinnitus near the low-frequency edge of the noise trauma<sup>17,20–22</sup>. However, few studies have directly compared neurophysiology and behavioural observations from the same animals<sup>20,23</sup>. It was therefore of great interest to us to relate noise-induced plasticity to perceptual disturbances.

Each of the eighteen noise-exposed rats used in this study was significantly impaired in its ability to detect a gap in narrowband noise centred on 8 or 10 kHz, but showed no impairment when the gap occurred in narrowband noise centred on 2 or 4 kHz or in broadband noise (Fig. 2, 4 weeks after exposure). Several studies have concluded that a frequency-specific impairment in gap detection is a likely sign that noise-exposed rats experience a mid-frequency tinnitus percept which fills the silent gaps<sup>17,23</sup> (Methods and Supplementary Figs 6–9). Although it is not possible to evaluate the subjective experience of rats definitively, the gap impairment has been taken as a possible behavioural correlate of tinnitus.

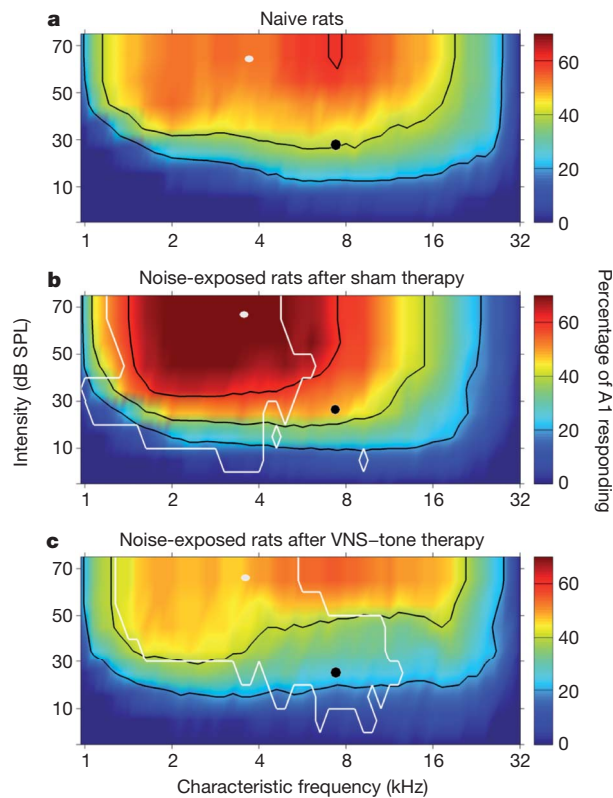
Map distortion and tuning curve broadening (but not changes in spontaneous activity or synchronization) were significantly correlated with the degree of gap impairment in untreated noise-exposed rats ( $R > 0.7$  (Pearson correlation coefficient),  $P < 0.05$ ,  $n = 8$  sham rats; Figs 3a, b and 4a–d and Supplementary Fig. 13). These correlations must be interpreted with caution because any variability in the initial cochlear trauma could generate a correlation between neural and behavioural changes even in the absence of a causal relationship. Though still not definitive, the best test for a causal relationship would be to reverse specifically the plasticity generated by noise exposure and document the reversal of the gap detection impairment.



**Figure 2 | VNS/multiple tone pairing eliminates the behavioural correlate of tinnitus.** Four weeks after noise exposure, each of the rats in both groups was unable to detect a gap in one or more of the narrowband noises tested ( $P > 0.05$ ; Supplementary Fig. 8b). The frequency with the greatest impairment four weeks after noise exposure is the putative tinnitus frequency for each rat. For both groups, gap detection at the putative tinnitus frequency was significantly impaired in comparison to broadband noise ( $P < 0.05$ ). The gap detection at the non-tinnitus frequency is based on gap detection in 16-kHz narrowband noise. **a**, Gap detection at the putative tinnitus frequency (dotted line) improved significantly after ten days of VNS-tone pairing, and the improvement persisted at least until the acute physiology experiment ( $n = 5$  rats). **b**, The sham group ( $n = 9$  rats) continued to be impaired. Two sham rats did not contribute data at the non-tinnitus frequency because they showed gap impairments at 16 kHz (as well as 8 and 10 kHz) four weeks after noise exposure. Black and grey horizontal bars represent duration of VNS and sham therapy, respectively. Asterisks represent significant differences ( $P < 0.05$ ) in gap detection at the putative tinnitus frequency between VNS therapy and sham therapy rats. Error bars, s.e.m.

We speculated that pairing VNS with randomly interleaved pure tones that span the rat hearing range, but exclude the overrepresented frequencies, could decrease the cortical representation of the excluded frequencies<sup>24</sup>. We also expected that pairing multiple tone frequencies with VNS ('VNS/multiple tone' pairing) would increase frequency selectivity and decrease synchronization as in our earlier nucleus basalis stimulation experiments<sup>25</sup>. We quantified behavioural and physiological correlates of tinnitus in noise-exposed rats and then tested whether pairing VNS with multiple tone frequencies could reverse the pathological plasticity and eliminate the perceptual disturbance in these rats.

VNS was repeatedly paired with multiple pure tones 300 times per day for 18 days in seven noise-exposed rats with impaired gap detection for mid-frequency sounds (Methods). Because we found that gap impairment occurred at 8–10 kHz, we selected the frequency of each randomly interleaved tone to be 1.3, 2.2, 3.7, 17.8 or 29.9 kHz. This pairing procedure was chosen because previous studies suggest it would reduce the cortical response to mid-frequency tones, increase frequency selectivity and decrease cortical synchronization<sup>2,25</sup>. After ten days of therapy, each of the seven rats showed a significant startle reduction in cued trials relative to uncued trials for every frequency tested ( $P < 0.05$ ; Fig. 2a and Supplementary Fig. 9a). Thus, pairing of VNS with multiple tones reversed the behavioural effect of noise exposure, which suggests that the rats' presumed tinnitus was no longer present. In contrast, rats in the sham therapy group showed a consistent

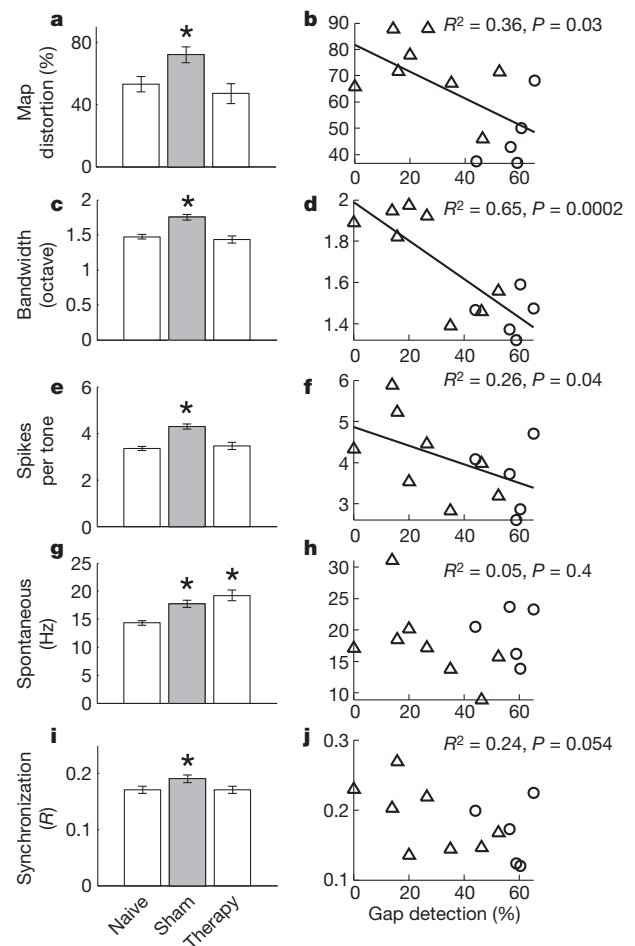


**Figure 3 | VNS/multiple tone pairing reverses map distortion.** The increased response of A1 neurons to tones following noise exposure is reversed by VNS/multiple tone pairing. **a**, Colour indicates the percentage of A1 neurons in naive rats that respond to a tone of any frequency and intensity combination. **b**, Percentage of A1 neurons that respond to each tone in noise-exposed rats that received sham therapy. **c**, Percentage of A1 neurons that respond to each tone in noise-exposed rats that received the VNS/multiple tone therapy. Black contour lines indicate 20, 40, and 60% responses. The white lines in **b** surround the regions of tones that are significantly increased ( $P < 0.01$ ) in comparison with naive rats. The white lines in **c** indicate significant decreases ( $P < 0.01$ ) in comparison with noise-exposed sham therapy rats. The filled white circles indicate the tone for which the increase in the number of cortical neurons was greatest, which is used to quantify the degree of map distortion in Fig. 4a, b. The filled black circles indicate the tone for which the proportional increase was greatest.

impairment in their ability to detect gaps in the putative tinnitus frequency (Fig. 2b). Each of the nine rats that received sham therapy (tones with no VNS, VNS with no tones or no therapy;  $n = 4, 2, 3$  rats, respectively) did not show a significant startle reduction in cued trials ( $P > 0.05$ ; Supplementary Fig. 9b) for at least one of the frequencies tested at each time point.

In the rats that received VNS paired with multiple tones, the impairment in gap detection was also eliminated when measured one day, one week and three weeks after the end of the therapy. This impairment was maintained in all three control groups at every time point tested (Supplementary Figs 9 and 10). These results indicate that pairing VNS with multiple tone frequencies is sufficient to eliminate the gap impairment induced by noise exposure (Supplementary Fig. 11). This is the first method reported to generate a long-lasting reversal of a behavioural correlate of chronic tinnitus.

Three weeks after the end of VNS/multiple tone pairing or sham therapy, we evaluated the physiological properties of the auditory cortex of each rat to determine whether the restored behaviour in the treated group was due to renormalization of the auditory cortex. After VNS/multiple tone pairing, most of the A1 properties that were degraded by noise exposure returned to normal levels. For example, the proportion of A1 neurons with characteristic frequencies between 12 and 23 kHz was indistinguishable from that in naive controls after VNS/multiple tone treatment (naive,  $20 \pm 2\%$ ; sham,  $15 \pm 5\%$ ; therapy,  $30 \pm 9\%$ ; Supplementary



**Figure 4 | Neurophysiological properties of naive, sham and therapy rats.**

**a, c, e, g, i**, Noise exposure caused a significant map distortion (**a**), decreased frequency selectivity (**c**), increased the tone-evoked response (**e**), increased the spontaneous rate (**g**) and increased the degree of cortical synchronization (**i**). VNS/multiple tone pairing returned each of these parameters, except spontaneous activity, to normal levels. **b, d, f**, Map organization (**b**), frequency selectivity (**d**) and tone-evoked response strength (**f**) were all correlated with the degree of gap impairment in individual rats. **h, j**, Spontaneous activity (**h**) and synchronization (**j**) were not significantly correlated with gap impairment. Each rat's gap detection ability was quantified as the average gap detection at the putative tinnitus frequency of each rat, averaged across the four time points collected after the beginning of therapy (Fig. 2). Error bars, s.e.m. Asterisks represent significant differences compared with naive rats ( $P$  values as indicated). Triangles and circles represent rats from the sham and therapy groups, respectively.

Figs 12 and 13a). The proportion of A1 neurons responding to 4-kHz, 70-dB SPL tones significantly increased relative to naive controls in sham rats and returned to normal levels in rats that had received the therapy three weeks earlier (naive,  $45.4 \pm 5.0\%$ ; sham,  $74.1 \pm 7.6\%$ ; therapy,  $49.1 \pm 6.6\%$ ; Figs 3 (white circles) and 4a). The degree of low-frequency map distortion was positively correlated with the degree of gap impairment observed in individual rats (Fig. 4b and Supplementary Fig. 13b). The percentage of cortex responding to 8-kHz, 30-dB SPL tones (Fig. 3, black circles) was also well correlated with the gap detection impairment ( $R^2 = 0.51$ ,  $P = 0.006$ ). These results support the earlier hypothesis that changes in cortical maps are causally related to tinnitus<sup>4,26</sup>.

VNS/multiple tone pairing reversed the increase in the width of frequency tuning of A1 multiunit activity (that is, decreased frequency selectivity) observed in noise-exposed rats (Fig. 4c). The bandwidth (measured at 10, 20, 30 or 40 dB above threshold) averaged across all A1 sites was highly correlated with the degree of gap impairment (Fig. 4d and Supplementary Fig. 14), thus supporting the earlier hypothesis that decreased frequency selectivity is causally related to tinnitus<sup>27</sup>.



VNS/multiple tone pairing reversed the increase in cortical excitability observed in noise-exposed rats (Fig. 4e). The average number of spikes evoked by tones within each site's receptive field was weakly correlated with the degree of impairment of gap detection (Fig. 4f), supporting the earlier hypothesis that tinnitus is related to increased excitability of cortical neurons<sup>28,29</sup>.

Finally, VNS/multiple tone pairing also reversed the increase in cortical synchronization observed in noise-exposed rats, but did not reverse the increase in cortical spontaneous activity observed in noise-exposed rats (Fig. 4g, i). There was a trend for the degree of synchronization to be correlated with the degree of gap impairment, but no correlation between the rate of spontaneous activity and the degree of gap impairment (Fig. 4h, j). Our observation that noise-induced increases in spontaneous activity and synchronization are not significantly correlated with behavioural correlates of tinnitus in individual rats is consistent with earlier reports<sup>19,23</sup>. However, given the potential for small changes in anaesthesia level to influence spontaneous activity and synchronization in the cortex, it remains a possibility that these factors contribute to tinnitus.

Hearing loss, hyperacusis and tinnitus often result from noise exposure and could contribute to the gap impairments observed in this study. Our results confirm that exposure to intense, high-frequency noise causes pathological plasticity that is well correlated with the inability to detect a gap in a mid-frequency, 65-dB SPL tone. Correlations alone do not suggest that these changes cause tinnitus because another confounding factor (such as variability in the degree of cochlear trauma) could cause both variables to be correlated without a causal connection. By randomizing the treatment of rats with identical noise exposure, we were able to eliminate the potential confound caused by variability in the response to noise exposure. Thus, our observation that pairing multiple tone frequencies with VNS can reverse both the neural and behavioural correlates of tinnitus provides good evidence that abnormal activity in the central auditory system is responsible for the subjective experience of tinnitus. In addition, neural correlates of hearing loss (tone thresholds) and hyperacusis (rate level functions) were not correlated with gap impairment in the rats tested (Supplementary Information). Thus, it is reasonable to conclude that the gap impairments observed in this study are primarily related to tinnitus.

VNS-directed plasticity represents a potentially powerful approach to treating tinnitus. Unlike pharmaceutical approaches, this method provides the possibility of generating long-lasting and stimulus-specific changes to neural circuits with minimal side effects. Our control experiments demonstrate that VNS-directed plasticity is driven by the repeated association of VNS with tones, and not by VNS alone. Additional studies are needed to determine whether the pairing of other sensory events with brief periods of VNS could be used to reverse the pathological plasticity associated with other common neurological conditions, such as chronic pain and amblyopia.

## METHODS SUMMARY

The VNS–tone pairing protocols, noise exposure procedure, gap detection testing, neurophysiology techniques and analysis are described in Methods. The noise exposure procedure, gap detection testing, and neurophysiology techniques were identical to those in earlier reports<sup>8,17,25</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 14 June; accepted 8 November 2010.**

**Published online 12 January 2011.**

1. Flor, H. Phantom-limb pain as a perceptual correlate of cortical reorganization following arm amputation. *Nature* **375**, 482–484 (1995).
2. Eggermont, J. J. & Roberts, L. E. The neuroscience of tinnitus. *Trends Neurosci.* **27**, 676–682 (2004).
3. Möller, A. R. Tinnitus and pain. *Prog. Brain Res.* **166**, 47–53 (2007).
4. Mühlhölzer, W., Elbert, T., Taub, E. & Flor, H. Reorganization of auditory cortex in tinnitus. *Proc. Natl Acad. Sci. USA* **95**, 10340–10343 (1998).

5. De Ridder, D., De Mulder, D., Menovsky, T., Sunaert, S. & Kovacs, S. Electrical stimulation of auditory and somatosensory cortices for treatment of tinnitus and pain. *Prog. Brain Res.* **166**, 377–388 (2007).
6. Okamoto, H., Stracke, H., Stoll, W. & Pantev, C. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proc. Natl Acad. Sci. USA* **107**, 1207–1210 (2010).
7. Flor, H., Denke, C., Schaefer, M. & Grüsser, S. Effect of sensory discrimination training on cortical reorganization and phantom limb pain. *Lancet* **357**, 1763–1764 (2001).
8. Kilgard, M. P. & Merzenich, M. M. Cortical map reorganization enabled by nucleus basalis activity. *Science* **279**, 1714–1718 (1998).
9. Dorr, A. E. & Debonnel, G. Effect of vagus nerve stimulation on serotonergic and noradrenergic transmission. *J. Pharmacol. Exp. Ther.* **318**, 890–898 (2006).
10. Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A. & Jensen, R. A. Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nature Neurosci.* **2**, 94–98 (1999).
11. Bao, S., Chan, V. T. & Merzenich, N. M. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* **412**, 79–83 (2001).
12. Bollinger, J. J. *Adult Auditory Cortical Plasticity Modulated by Locus Coeruleus Activity*. PhD thesis, Univ. California San Francisco (2006).
13. Ben-Menachem, E. Vagus nerve stimulation, side effects, and long-term safety. *J. Clin. Neurophysiol.* **18**, 415–418 (2001).
14. Noreña, A. J., Tomita, M. & Eggermont, J. J. Neural changes in cat auditory cortex after a transient pure-tone trauma. *J. Neurophysiol.* **90**, 2387–2401 (2003).
15. Salvi, R. J., Wang, J. & Ding, D. Auditory plasticity and hyperactivity following cochlear damage. *Hear. Res.* **147**, 261–274 (2000).
16. Eggermont, J. in *Tinnitus: Pathophysiology and Treatment* (eds Langguth, B., Hajak, G., Kleinjung, T., Cacace, A. & Möller, A. R.) 19–35 (Prog. Brain Res. 166, Elsevier, 2007).
17. Turner, J. G. et al. Gap detection deficits in rats with tinnitus: a potential novel screening tool. *Behav. Neurosci.* **120**, 188–195 (2006).
18. Murphy, W. J. & van Campen, L. E. Temporary threshold shift in ABRs and DPOAEs following noise exposure in Long–Evans rats. *J. Acoust. Soc. Am.* **109**, 2373 (2001).
19. Bauer, C. A., Turner, J. G., Caspary, D. M., Myers, K. S. & Brozoski, T. J. Tinnitus and inferior colliculus activity in chinchillas related to three distinct patterns of cochlear trauma. *J. Neurosci. Res.* **86**, 2564–2578 (2008).
20. Bauer, C. A. & Brozoski, T. J. Assessing tinnitus and prospective tinnitus therapeutics using a psychophysical animal model. *J. Assoc. Res. Otolaryngol.* **2**, 54–64 (2001).
21. Lobarinas, E., Sun, W., Cushing, R. & Salvi, R. A novel behavioral paradigm for assessing tinnitus using schedule-induced polydipsia avoidance conditioning (SIP-AC). *Hear. Res.* **190**, 109–114 (2004).
22. Moore, B. C. & Sandhya, V. The relationship between tinnitus pitch and the edge frequency of the audiogram in individuals with hearing impairment and tonal tinnitus. *Hear. Res.* **261**, 51–56 (2010).
23. Yang, G. et al. Salicylate induced tinnitus: behavioral measures and neural activity in auditory cortex of awake rats. *Hear. Res.* **226**, 244–253 (2007).
24. Noreña, A. J. & Eggermont, J. J. Enriched acoustic environment after noise trauma reduces hearing loss and prevents cortical map reorganization. *J. Neurosci.* **25**, 699–705 (2005).
25. Kilgard, M. P., Vazquez, J. L., Engineer, N. D. & Pandya, P. K. Experience dependent plasticity alters cortical synchronization. *Hear. Res.* **229**, 171–179 (2007).
26. Dietrich, V., Nieschalk, M., Stoll, W., Rajan, R. & Pantev, C. Cortical reorganization in patients with high frequency cochlear hearing loss. *Hear. Res.* **158**, 95–101 (2001).
27. Dauman, R. & Cazals, Y. Auditory frequency selectivity and tinnitus. *Arch. Otorhinolaryngol.* **246**, 252–255 (1989).
28. Kaltenbach, J. A. & Afman, C. E. Hyperactivity in the dorsal cochlear nucleus after intense sound exposure and its resemblance to tone-evoked activity: a physiological model for tinnitus. *Hear. Res.* **140**, 165–172 (2000).
29. Diesch, E., Andermann, M., Flor, H. & Rupp, A. Interaction among the components of multiple auditory steady-state responses: enhancement in tinnitus patients, inhibition in controls. *Neuroscience* **167**, 540–553 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank A. Kuzu, J. Omana, D. Vuppala, H. Rasul, M. Fink, E. Hanack, R. Miller and C. Walker for help with rat behavioural training. We would also like to thank J. Eggermont, A. Möller, C. Bauer, J. Fritz, H. Reed, C. Engineer, A. Reed, M. Brosch, R. Rennaker, R. Beitel, V. Miller, C. McIntyre, G. White, P. Pandya, R. Tyler and D. deRidder for suggestions about earlier versions of the manuscript. This work was supported by the James S. McDonnell Foundation, the Texas Advanced Research Program, the National Institute for Deafness and other Communication Disorders, and MicroTransponder Inc.

**Author Contributions** N.D.E., J.R.R., J.D.S., S.P.S. and M.S.B. did the behaviour training sessions, noise exposure and auditory brainstem response recordings. N.D.E., J.R.R., J.D.S., W.A.V. and J.A.S. did cortical microelectrode mappings. J.A.S. did the A1 mapping surgeries. S.P.S. and N.D.E. did all the VNS implant surgeries. M.P.K. and N.D.E. designed the experiments, wrote the manuscript and performed data analysis. All authors discussed the paper and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.D.E. ([navez@utdallas.edu](mailto:navez@utdallas.edu)).

## METHODS

**VNS surgical protocol.** Female Sprague–Dawley rats (250–350 g) were implanted with a platinum–iridium bipolar cuff electrode around the left cervical vagus nerve<sup>9</sup>. As in humans, only the left vagus nerve was stimulated because the right vagus nerve contains efferents that stimulate the sinoatrial node and can cause cardiac complications<sup>13</sup>. Leads from the electrode were tunnelled subcutaneously to the top of the head. A four-channel connector was used to deliver current to the stimulating electrode and monitor the electroencephalogram (EEG) during daily VNS sessions. Bone screws placed over the vertex and the cerebellum were used to record auditory brainstem responses (ABRs) and EEG. Each rat was given antibiotics to prevent infection and a single dose of atropine and dexamethazone to reduce fluid accumulations in the lungs immediately after completion of the surgery.

**VNS stimulation parameters and single-tone pairing procedures.** VNS was delivered to unanaesthetized, unrestrained rats in a  $25 \times 25 \times 25$  cm<sup>3</sup> wire cage, located inside a  $50 \times 60 \times 70$  cm<sup>3</sup> chamber lined with acoustic insulating foam. A pilot study was conducted to determine the minimal VNS parameters that reliably reduced EEG amplitude during slow-wave sleep (Supplementary Fig. 3). VNS parameters were identical for every rat in this study. Each 100- $\mu$ s, charge-balanced biphasic pulse was delivered with a current of 0.8 mA. The stimulation was delivered as a train of 15 pulses at 30 Hz (500-ms train duration). Cuff impedances were measured daily ( $\sim 5$  k $\Omega$ ). The impedance for three rats was unusually high after implantation and these rats were assigned to the tone-alone and no-therapy groups. The impedance was stable across the duration of training for all other rats. The 500-ms pure tones began 150 ms after the onset of the VNS train (Supplementary Figs 1 and 2). For our earlier nucleus basalis stimulation studies, stimulation beginning either 200 ms before tone onset or 50 ms after tone onset generated indistinguishable map plasticity<sup>8</sup>.

VNS was delivered 300 times per day for 20 days, during a VNS–tone pairing session that lasted 2.5 h (Supplementary Figs 2). To prevent rats from anticipating stimulation timing, there was a 50% chance that VNS would be delivered every 15 s. Twenty-four hours after the last pairing, rats were anaesthetized with pentobarbital and the right auditory cortex was exposed to allow for high-density extracellular microelectrode mapping<sup>8</sup>.

One group of rats ( $n = 8$ ) was exposed to a single 9-kHz, 60-dB SPL tone paired with VNS. No sound was presented when VNS was not delivered. A second group ( $n = 5$ ) was exposed to a 19-kHz, 50-dB SPL tone paired with VNS. During the trials in which no VNS was delivered (50%), a 4-kHz, 50-dB SPL tone was presented. As a result, a 19- or 4-kHz tone was delivered every 15 s. Frequency and intensity calibrations were performed with an ACO Pacific microphone (PS9200-7016) and Tucker-Davis Technologies SIGCAL v4.2 software. The free-field tones were presented from a speaker (Optimus) suspended 20 cm above the wire cage. All paired tones had a 5-ms rise–fall time. The intensity of every tone was selected to be approximately 20 dB SPL above the rat hearing threshold.

**Noise exposure and ABRs.** Twenty-eight experimental and control rats were barbiturate-anaesthetized and exposed to 16-kHz, 115-dB SPL, octave-band noise for 1 h (refs 17, 20). A single speaker was positioned 5 cm from the left ear. No ear plugs were used to restrict the noise exposure to one ear. Bilateral noise exposure was used because it best approximates the noise exposure that occurs in humans. To confirm cochlear trauma, elevated thresholds were quantified using ABRs in ten rats under pentobarbital anaesthesia before noise exposure, immediately after exposure and 11 weeks after noise exposure (when the auditory cortex was mapped). For ABR recordings, the speaker was positioned 10 cm from the left ear and pure tones (10 ms long, 2.5-ms rise–fall time) were delivered at a rate of 20 Hz. Tone frequencies were 4, 10, 16 and 32 kHz in 10-dB steps from 0 to 85 dB SPL. Tones were randomly interleaved with 1,500 repeats for each frequency–intensity combination. The signals were filtered from 100 to 3,000 Hz and recorded using BRAINWARE v8.12 (Tucker-Davis Technologies). Threshold was defined as the lowest 10-dB SPL step at which an ABR could be recognized (Supplementary Fig. 4).

**Gap detection testing.** The Turner gap detection method was used to assess a behavioural correlate of tinnitus in every noise-exposed rat<sup>17</sup> (Supplementary Figs 6–8). This method has previously been cross-validated with a conditioned lever suppression task<sup>20</sup> ( $R = 0.75$ ) and a licking suppression task<sup>21</sup>. The gap detection method was selected because it avoids the need for food or water deprivation, electric shock or months of behavioural training<sup>17</sup>. Testing took place in a  $20 \times 20 \times 20$  cm<sup>3</sup> wire-mesh cage in a  $67 \times 67 \times 67$  cm<sup>3</sup> chamber lined with 5-cm acoustic foam. The cage was placed on a startle platform (Lafayette Instrument Co.) that used a piezoelectric transducer to generate a continuous record of downward force. Sounds were generated using System 3 hardware and software (Tucker-Davis Technologies) and were delivered by a speaker (Tucker-Davis Technologies FF1) mounted 20 cm above the cage. Rats underwent gap detection testing with different band-pass-filtered (1,000-Hz bandwidth) sounds centred at 2, 4, 8, 10, 16, 20 and 24 kHz at 65 dB SPL (ref. 17). Startle responses were elicited by a 20-ms burst of white noise at 100 dB SPL. In 50% of trials, a 50-ms gap embedded in the continuous sound served as a warning of a subsequent startling noise and allowed

rats to reduce the amplitude of the response (Supplementary Fig. 7b). The gap in the narrowband noise began 100 ms before the onset of the broadband startling noise. Rats underwent 30 trials during each session. The order of sessions with different continuous sounds was counterbalanced across rats. The interval between each startle sound was 30–35 s.

In untreated noise-exposed rats, gaps in a specific narrowband sound (usually 8 or 10 kHz) did not serve as an effective warning, presumably because the ongoing tinnitus percept prevented the rats from detecting the silent gap. Thus, the animals were not warned that a loud startling noise was coming and exhibited a strong startle response (Supplementary Figs 7b and 8b). Gap detection was quantified as one minus the ratio of the startle amplitude when the startling noise was preceded by a gap in the 65-dB SPL, continuous narrowband sound to the startle amplitude when the startling noise was not preceded by a warning gap. Supplementary Fig. 8 shows typical data from one noise-exposed rat for a session in which the noise burst was cued with a gap in broadband noise (left) and a session in which a gap in an 8-kHz tone served as the warning cue (right). The warning gap typically reduced the startle amplitude by 60–70% (Supplementary Fig. 8a). In noise-exposed rats, gaps in the narrowband noise centred near the low edge of the trauma noise typically reduced the startle amplitude by less than 20%, which is not a statistically significant reduction (Supplementary Fig. 8b). The same procedure was also administered using gaps in 65-dB SPL broadband noise as warning cues of the startling noise (Supplementary Fig. 7a). The frequency with the greatest impairment four weeks after noise exposure is the putative tinnitus frequency for each rat (Fig. 2).

Thirty-six rats were initially tested using the gap startle task for inclusion in this study. Five rats were excluded from the study because they showed no detectable startle response to the noise burst. Of the 31 remaining rats, three were excluded because their startle responses were unusually variable. Twenty-eight rats received noise exposure. Eighteen of these showed a statistically significant impairment in the detection of gaps in one or both mid-frequency (8- or 10-kHz) narrowband sounds tested, relative to gap detection before noise exposure ( $P < 0.05$ ). Three rats were excluded from further study because they no longer showed a startle response to the noise burst (that is, could no longer detect the startle stimulus). Seven rats were excluded from further study because they showed no impairment in gap detection (that is, no evidence of tinnitus). Our observation that gap impairments do not always result from noise exposure is consistent with human and animal studies showing that although hearing loss is common in individuals with tinnitus, the majority of individuals with hearing loss do not have tinnitus<sup>20,30,31</sup>.

Each of the eighteen rats included in this study showed a significant impairment in its ability to detect a gap in narrowband noise centred on 8 kHz (16 of 18) or 10 kHz (12 of 18). None of the 18 rats showed a significant impairment in the ability to detect a gap in low-frequency narrowband noises (2 or 4 kHz) or in broadband noise (Fig. 2 and Supplementary Fig. 11). This result indicates that these rats are able to respond normally to the startling noise burst and that the mechanisms for modulating the startle response using silent gaps remain intact. Our observation that noise-exposed rats can show gap detection impairments centred at a single frequency or across a narrow range of frequencies is consistent with clinical studies showing significant heterogeneity across subjects in the spectral characteristics of the tinnitus percept<sup>22,32,33</sup>. Despite this heterogeneity, a large fraction of tinnitus patients can match their tinnitus to a pitch and describe their phantom sound as tonal<sup>22</sup>.

**VNS tone delivery to noise exposed rats.** Rats were tested for gap impairment four weeks after noise exposure and 10 and 20 days after the beginning of the sham or experimental therapy. In the VNS/multiple tone paired group ( $n = 5$  rats), tones were paired with VNS every 15 s with no VNS–tone pairing 50% of the time. The tone frequencies paired with VNS in the therapy group were designed to reduce the 8–10-kHz region of the frequency map. VNS was repeatedly paired with a 1.3-, 2.2-, 3.7-, 17.8- or 29.9-kHz tone that was randomly selected every trial (300 trials per day). Each tone was presented at  $\sim 20$  dB above the normal hearing threshold for that frequency. The tone-alone control group was passively exposed to the same tones on the same schedule as used in the paired group. A VNS-alone control group received VNS stimulation on the same schedule as used in the paired group without presentation of tones. The third control group did not receive tones or VNS.

To test whether the tinnitus percept remained suppressed after the end of VNS–tone pairing, rats were also tested on gap detection one and three weeks after the end of therapy. At the end of three weeks (that is, 11 weeks after noise exposure), multiunit responses were recorded from auditory cortex neurons from the therapy, sham and naive control rats using dense microelectrode mapping techniques. Physiological and behavioural results from the tone-alone, VNS-alone and no-therapy groups were statistically indistinguishable (Supplementary Fig. 10 and physiological data not shown). Data from the three groups are combined and referred to as sham controls in the main text.

**Neurophysiology.** In this study, we recorded from a total of 1,492 sites in 21 rats ( $n = 8$  naive controls,  $n = 5$  VNS therapy and  $n = 8$  sham controls). Nine hundred and sixty-five of those sites were in A1 and were included in the analysis presented in this report. We recorded 220 multiunit responses from A1 sites in noise-exposed rats that received VNS/multiple tone pairing ( $n = 5$ ). We also recorded 321 A1 sites from noise-exposed rats that did not receive VNS/tone pairing ( $n = 8$ ). The latter group included noise-exposed rats that received tones with no VNS ( $n = 3$ ), VNS with no tones ( $n = 2$ ) or no therapy ( $n = 3$ ). Because neural and behavioural responses were similar in all three control groups, the results were pooled to form a single data set referred to as the sham therapy group. During the acute electrophysiology recordings, sounds were delivered in a foam-lined, double-walled, sound-attenuated chamber using a speaker (Motorola 40-1221) positioned directly opposite the left ear at a distance of 10 cm. Multiunit responses were recorded using Parylene-coated tungsten electrodes that were glued together (250- $\mu\text{m}$  separation, 2 M $\Omega$  at 1 kHz; FHC) and lowered approximately 500  $\mu\text{m}$  below the cortical surface. Frequency and intensity calibrations were performed with an ACO Pacific microphone (PS9200-7016) and Tucker-Davis Technologies SIGCAL v4.2 software. Auditory frequency tuning curves were determined at each site by presenting 81 logarithmically spaced frequencies spanning 1 to 32 kHz at 16 intensities from 0 to 75 dB SPL (1,296 total stimuli). The tones (25-ms duration, 5-ms rise–fall time) were randomly interleaved and separated by 500 ms. Tuning curve parameters were determined by an experienced blind observer using custom software written in MATLAB v7.9 (Mathworks) to randomize the order of data from each recording site across all groups. Experimenters were blind to the experimental conditions of each rat during electrophysiology recordings.

**Data analysis.** Gap discrimination was quantified as the percentage inhibition of the startle response when a gap (warning cue) preceded the startling noise relative to the startle response when no gap was present<sup>17</sup>. Eight of 36 rats tested failed to generate consistent startle responses and were excluded from the study before noise exposure. Noise exposure eliminated the startle response in three of the remaining 28 rats, and these rats were excluded from the study. Noise exposure failed to generate any impairment in gap detection in seven of the remaining 25 rats, and these rats were also excluded from the study. Eighteen noise-exposed rats were included in the study. Neural responses were collected from thirteen rats (five

VNS–tone paired rats and eight sham therapy rats). One rat died before neural responses could be collected. Only behavioural responses (and EEG) were collected from the remaining four rats (two treated rats and two shams) so that the duration of the benefit could be estimated.

Sites were determined to be in A1 on the basis of continuous tonotopy. At each A1 recording site, characteristic frequency, frequency bandwidth, response threshold, spontaneous rate and latency were determined using a standard method in which the experimenter was blind to the experimental group and recording location<sup>8</sup>. At each pair of simultaneously recorded A1 sites, neural synchrony during silence (300 s) was quantified as the cross-correlation function<sup>25</sup>. The peak in the cross-correlation function (with or without subtraction of the shift predictor) was also computed and gave similar results to Pearson correlation coefficient ( $R$ ). Map plasticity was quantified as the percent of A1 neurons with a characteristic frequency in a given range or as the percent of A1 neurons responding to each frequency–intensity combination using the Voronoi tessellation method of interpolation<sup>8,34</sup>. Frequency selectivity was quantified as the bandwidth 10, 20, 30 or 40 dB above threshold. Results were similar regardless of the intensity above threshold used. Excitability was quantified as the number of spikes evoked by each tone within each site's receptive field and as the spontaneous activity rate during silence.

All protocols and recording procedures comply with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Animal Care and Use Committee at the University of Texas at Dallas.

30. Bauer, C. A., Brozoski, T. J. & Myers, K. Primary afferent dendrite degeneration as a cause of tinnitus. *J. Neurosci. Res.* **85**, 1489–1498 (2007).
31. König, O., Schaette, R., Kempter, R. & Gross, M. Course of hearing loss and occurrence of tinnitus. *Hear. Res.* **221**, 59–64 (2006).
32. Burns, E. M. A comparison of variability among measurements of subjective tinnitus and objective stimuli. *Audiology* **23**, 426–440 (1984).
33. Ochi, K., Ohashi, T. & Kenmochi, M. Hearing impairment and tinnitus pitch in patients with unilateral tinnitus: comparison of sudden hearing loss and chronic tinnitus. *Laryngoscope* **113**, 427–431 (2003).
34. Engineer, N. D. et al. Environmental enrichment improves response strength, threshold, selectivity, and latency of auditory cortex neurons. *J. Neurophysiol.* **92**, 73–82 (2004).



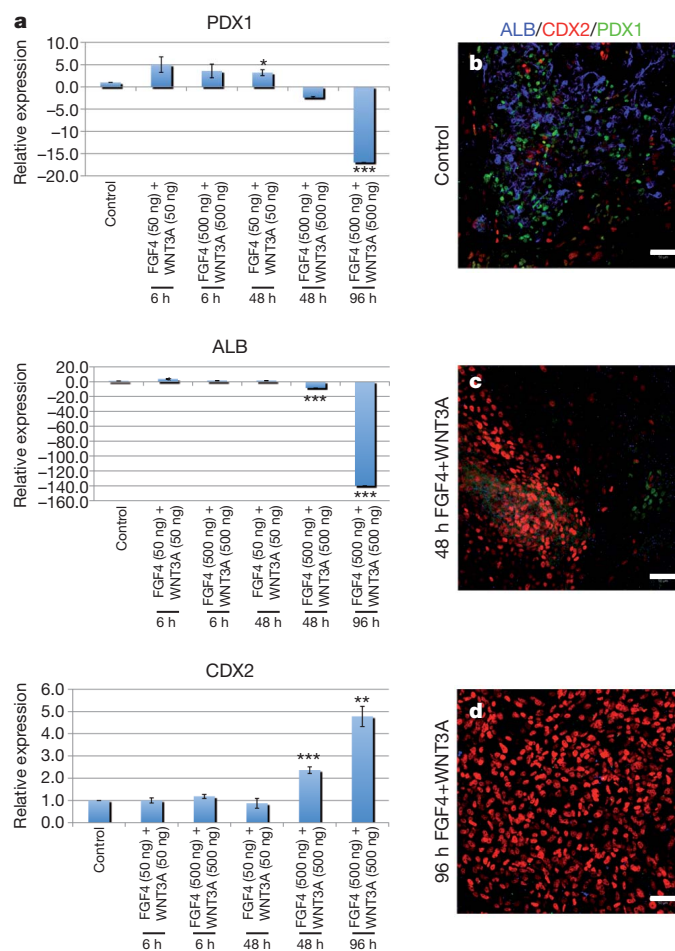
# Directed differentiation of human pluripotent stem cells into intestinal tissue *in vitro*

Jason R. Spence<sup>1</sup>, Christopher N. Mayhew<sup>1</sup>, Scott A. Rankin<sup>1</sup>, Matthew F. Kuhar<sup>1</sup>, Jefferson E. Vallance<sup>2</sup>, Kathryn Tolle<sup>1</sup>, Elizabeth E. Hoskins<sup>3</sup>, Vladimir V. Kalinichenko<sup>1,4</sup>, Susanne I. Wells<sup>3</sup>, Aaron M. Zorn<sup>1</sup>, Noah F. Shroyer<sup>1,2</sup> & James M. Wells<sup>1</sup>

Studies in embryonic development have guided successful efforts to direct the differentiation of human embryonic and induced pluripotent stem cells (PSCs) into specific organ cell types *in vitro*<sup>1,2</sup>. For example, human PSCs have been differentiated into monolayer cultures of liver hepatocytes and pancreatic endocrine cells<sup>3–6</sup> that have therapeutic efficacy in animal models of liver disease<sup>7,8</sup> and diabetes<sup>9</sup>, respectively. However, the generation of complex three-dimensional organ tissues *in vitro* remains a major challenge for translational studies. Here we establish a robust and efficient process to direct the differentiation of human PSCs into intestinal tissue *in vitro* using a temporal series of growth factor manipulations to mimic embryonic intestinal development<sup>10</sup>. This involved activin-induced definitive endoderm formation<sup>11</sup>, FGF/Wnt-induced posterior endoderm patterning, hindgut specification and morphogenesis<sup>12–14</sup>, and a pro-intestinal culture system<sup>15,16</sup> to promote intestinal growth, morphogenesis and cytodifferentiation. The resulting three-dimensional intestinal ‘organoids’ consisted of a polarized, columnar epithelium that was patterned into villus-like structures and crypt-like proliferative zones that expressed intestinal stem cell markers<sup>17</sup>. The epithelium contained functional enterocytes, as well as goblet, Paneth and enteroendocrine cells. Using this culture system as a model to study human intestinal development, we identified that the combined activity of WNT3A and FGF4 is required for hindgut specification whereas FGF4 alone is sufficient to promote hindgut morphogenesis. Our data indicate that human intestinal stem cells form *de novo* during development. We also determined that NEUROG3, a pro-endocrine transcription factor that is mutated in enteric anendocrinosis<sup>18</sup>, is both necessary and sufficient for human enteroendocrine cell development *in vitro*. PSC-derived human intestinal tissue should allow for unprecedented studies of human intestinal development and disease.

The epithelium of the intestine is derived from a simple sheet of cells called the definitive endoderm<sup>17</sup>. As a first step to generating intestinal tissue from PSCs (summarized in Supplementary Fig. 1), we used activin A, a nodal-related TGF- $\beta$  molecule, to promote differentiation into definitive endoderm as previously described<sup>11</sup>, resulting in up to 90% of the cells co-expressing the definitive endoderm markers SOX17 and FOXA2 and fewer than 2% expressing the mesoderm marker brachyury (Supplementary Fig. 2a). Using microarray analysis we observed a robust activation of definitive endoderm markers, many of which were expressed in mouse definitive endoderm from embryonic day (e)7.5 embryos (Supplementary Fig. 3 and Supplementary Table 1a, b). We investigated the intrinsic ability of definitive endoderm to form foregut and hindgut lineages by culturing for 7 days under permissive conditions and observed that cultures treated with activin A for only 3 days were competent to develop into both foregut (albumin (ALB)<sup>+</sup> and PDX1<sup>+</sup>) and hindgut (CDX2) lineages (Fig. 1b, control). In contrast, treatment with activin A for 4–5 days resulted in

definitive endoderm cultures that were intrinsically anterior in character and less competent in forming posterior lineages (Supplementary Fig. 2b).

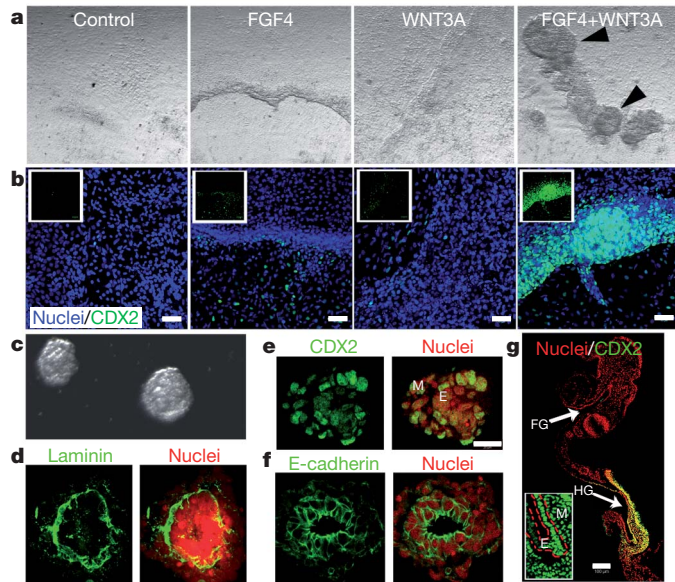


**Figure 1** | FGF4 and WNT3A act synergistically in a temporal and dose-dependent manner to specify stable posterior endoderm fate. **a–d**, Activin A (100 ng ml<sup>-1</sup>) was used to differentiate H9 human ES cells into definitive endoderm. Definitive endoderm was treated with the posteriorizing factors FGF4 (50 or 500 ng), WNT3A (50 or 500 ng), or both for 6, 48 or 96 h. Cells were placed in permissive media for 7 days and expression of foregut markers (ALB, PDX1) and the hindgut marker (CDX2) were analysed by RT-qPCR (**a**) and immunofluorescence (**b–d**). The definitive endoderm of controls was grown for identical lengths of time in the absence of FGF4 or WNT3A. High levels of FGF4+WNT3A for 96 h resulted in stable CDX2 expression and lack of foregut marker expression. Scale bars, 50  $\mu$ m. Error bars are s.e.m. ( $n = 3$ ). \* $P < 0.05$ , \*\* $P < 0.001$ , \*\*\* $P < 0.0001$ .

<sup>1</sup>Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA. <sup>2</sup>Division of Gastroenterology, Hepatology and Nutrition, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA. <sup>3</sup>Division of Hematology and Oncology Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA. <sup>4</sup>Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, Ohio 45229-3039, USA.

Having identified the window of time when definitive endoderm fate was plastic (day 3 of activin A treatment), we used WNT3A and FGF4 to promote hindgut and intestinal specification. Studies in mouse, chick and frog embryos have demonstrated that Wnt and FGF signalling pathways are required for repressing anterior development and promoting posterior endoderm formation into the midgut and hindgut<sup>12–14</sup>. Consistent with this, conditioned media containing WNT3A was recently shown to promote *Cdx2* expression in mouse embryonic stem (ES)-cell-derived embryoid bodies<sup>19</sup>. In human definitive endoderm cultures, neither factor alone was sufficient to robustly promote a posterior fate (Supplementary Fig. 2c); but high concentrations of both FGF4 and WNT3A (FGF4+WNT3A) induced expression of the hindgut marker CDX2 in the definitive endoderm after 48 h (Supplementary Fig. 4). However, 48 h of FGF4+WNT3A treatment did not stably induce a CDX2<sup>+</sup> hindgut fate and expression of anterior markers PDX1 and albumin reappeared after cells were cultured in permissive media for 7 days (Fig. 1a, c). In contrast, 96 h of exposure to FGF4+WNT3A resulted in stable CDX2 expression and absence of anterior markers (Fig. 1a, d). These findings indicate a previously unidentified requirement for the synergistic activities of both the FGF and Wnt pathways in specifying the CDX2<sup>+</sup> mid/hindgut lineage.

Remarkably, FGF4+WNT3A-treated cultures underwent morphogenesis that was similar to embryonic hindgut formation. Between 2 and 5 days of FGF4+WNT3A treatment, flat cell sheets condensed into CDX2<sup>+</sup> epithelial tubes, many of which budded off to form floating hindgut spheroids (Fig. 2a–c, Supplementary Fig. 5a–f and Supplementary Table 2a). Spheroids were similar to e8.5 mouse hindgut and consisted of uniformly CDX2<sup>+</sup> polarized epithelium surrounded by CDX2<sup>+</sup> mesenchyme (Fig. 2d–g). Spheroids were completely devoid



**Figure 2 | Morphogenesis of posterior endoderm into three-dimensional, hindgut-like spheroids.** **a**, Bright-field images of definitive endoderm cultured for 96 h in media, FGF4, WNT3A or FGF4+WNT3A. FGF4+WNT3A cultures contained three-dimensional epithelial tubes and free-floating spheres (black arrows) **b**, CDX2 immunostaining (green) and nuclear stain (DRAQ5, blue) on cultures shown in **a**. Insets show CDX2 staining alone. **c**, Bright-field image of hindgut-like spheroids. **a–c**, Scale bars, 50  $\mu$ m. **d–f**, Analysis of CDX2, basal-lateral laminin and E-cadherin expression demonstrates an inner layer of polarized, cuboidal, CDX2<sup>+</sup> epithelium surrounded by non-polarized mesenchymal CDX2<sup>+</sup> cells. Scale bar in **e** is 20  $\mu$ m. **g**, CDX2 expression in an e8.5 mouse embryo (sagittal section). Inset is a magnified view showing that both hindgut endoderm (E, outlined with a red dashed line) and adjacent mesenchyme (M) are CDX2 positive (green). FG, foregut; HG, hindgut. Scale bar, 100  $\mu$ m.

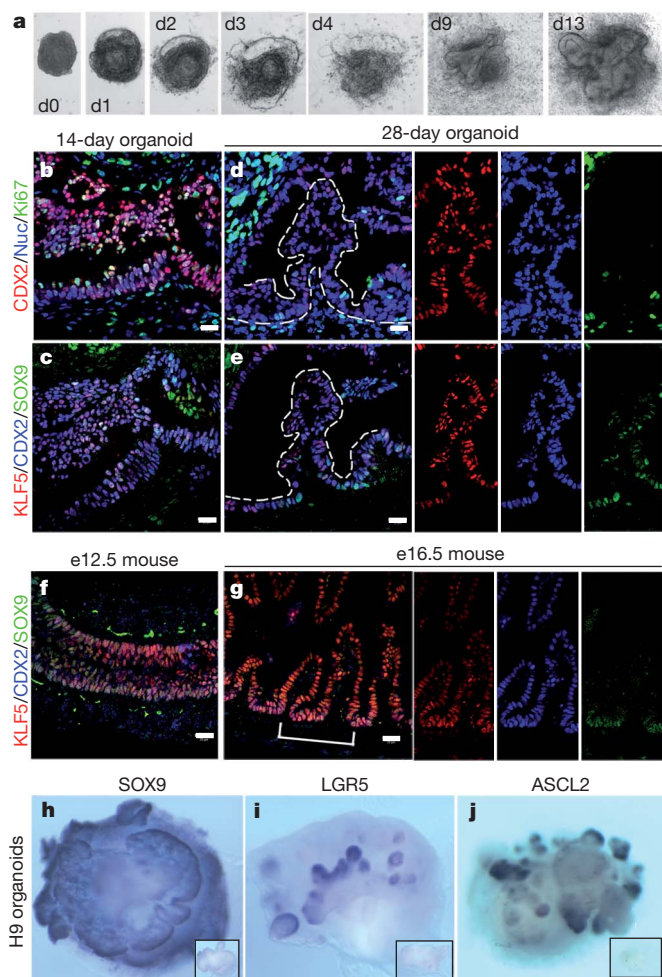
of albumin and PDX1-expressing foregut cells (Supplementary Fig. 5h, i). *In vitro* gut-tube morphogenesis was never observed in control or WNT3A-only treated cultures. FGF4-treated cultures had a twofold expansion of mesoderm and generated 4–10-fold fewer spheroids (Supplementary Fig. 2c and Supplementary Table 2a), which were weakly CDX2<sup>+</sup> and did not undergo further expansion (data not shown). Together our data support a mechanism for hindgut development where FGF4 promotes mesoderm expansion and morphogenesis, whereas FGF4 and WNT3A synergy is required for the specification of the hindgut lineage.

Importantly, this method for directed differentiation is broadly applicable to other PSC lines, as we were able to generate hindgut spheroids from both H1 and H9 human ES cell lines and from four induced PSC (iPSC) lines that we have generated and characterized (Supplementary Figs 3, 5 and 6). The kinetics of differentiation and the formation of spheroids were comparable between these lines (Supplementary Table 2). Two other iPSC lines tested were poor at hindgut spheroid formation and line iPSC3.6 also had a divergent transcriptional profile during definitive endoderm formation (Supplementary Fig. 3 and Supplementary Table 2c).

Whereas *in vivo* engraftment of PSC-derived cell types, such as pancreatic endocrine cells, has been used to promote maturation<sup>9</sup>, efficient development and maturation of organ tissues *in vitro* has proven more difficult. We investigated whether hindgut spheroids could develop and mature into intestinal tissue *in vitro* using recently described three-dimensional culture conditions that support growth and renewal of the adult intestinal epithelium<sup>15,16</sup>. When placed into this culture system, hindgut spheroids developed into intestinal organoids in a staged manner that was notably similar to fetal gut development (Fig. 3, Supplementary Fig. 5g and Supplementary Fig. 7). In the first 14 days the simple cuboidal epithelium of the spheroid expanded and formed a highly convoluted pseudostratified epithelium surrounded by mesenchymal cells (Fig. 3a–c), similar to an e12.5 fetal mouse gut (Fig. 3f). After 28 days, the epithelium matured into a columnar epithelium with villus-like involutions that protrude into the lumen of the organoid (Fig. 3d, e). Comparable transitions were observed during mouse fetal intestinal development (Fig. 3f, g and Supplementary Fig. 7). The spheroids expanded up to 40 fold in mass as they formed organoids (data not shown) and were split and passed over 9 additional times and cultured for over 140 days with no signs of growth failure. The cellular gain during that time was up to 1,800 fold (data not shown), resulting in a total cellular expansion of 72,000 fold per hindgut spheroid. This directed differentiation was up to 50 fold more efficient than spontaneous embryoid body differentiation methods<sup>20</sup> (Supplementary Fig. 8) and resulted in organoids that were almost entirely intestinal (Supplementary Fig. 2e–g) as compared to embryoid bodies that contained a mix of neural, vascular and epidermal tissues (Supplementary Fig. 8).

Marker analysis showed that after 14 days in culture, virtually all of the epithelium expressed the intestinal transcription factors CDX2, KLF5 and SOX9 broadly and was highly proliferative (Fig. 3b, c). By 28 days, CDX2 and KLF5 remained broadly expressed in over 90% of the epithelium (Supplementary Fig. 2), whereas SOX9 became localized to pockets of proliferating cells at the base of the villus-like protrusions (Fig. 3d, e) similar to the intervillus epithelium of fetal mouse intestines at e16.5 (Fig. 3g and Supplementary Fig. 9). 5-bromodeoxyuridine (BrdU) pulse chase and analysis of organoids using a Z-stack series of confocal microscopic images showed that epithelial BrdU incorporation was largely restricted to SOX9-expressing cells in crypt-like structures that penetrated into the underlying mesenchyme (Supplementary Fig. 9). At 28 days, LGR5 is not expressed and ASCL2 (ref. 21) is broadly expressed and not restricted to the SOX9<sup>+</sup> proliferative zone. However, organoids cultured until 56 days expressed both ASCL2 and LGR5 in restricted epithelial domains that appear to overlap with the SOX9<sup>+</sup> zone (Fig. 3h–j and Supplementary Fig. 10). This domain is similar to developing intestinal progenitor domains *in vivo*,

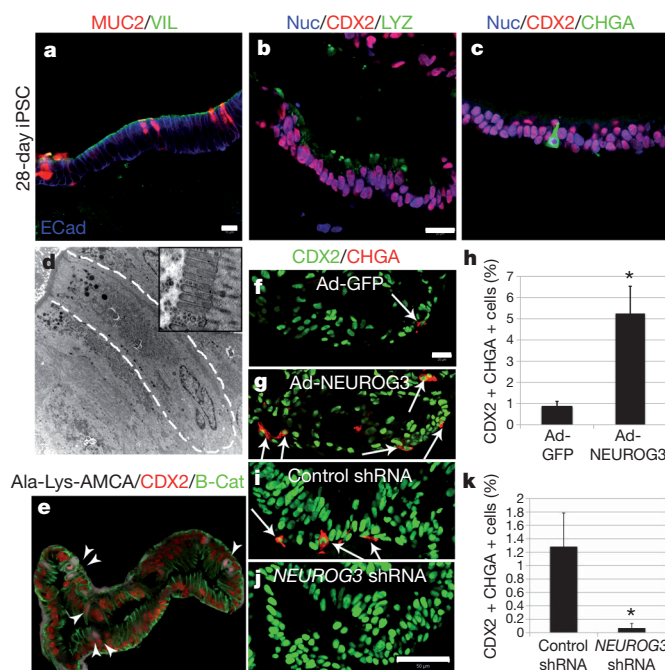




**Figure 3 | Human ES cells and iPSCs form three-dimensional intestine-like organoids.** **a**, A time course shows that intestinal organoids formed highly convoluted epithelial structures surrounded by mesenchyme after 13 days (**d**). **b–e**, Intestinal transcription factor expression (KLF5, CDX2, SOX9) and cell proliferation on serial sections of organoids after 14 and 28 days (serial sections are **b** and **c**, **d** and **e**). Ki67, nuclear proliferation antigen. Nuc, nuclei. **f, g**, Expression of KLF5, CDX2, and SOX9 in mouse fetal intestine at e14.5 (**f**) and e16.5 (**g**) is similar to developing intestinal organoids. The right panels show separate colour channels for **d, e** and **g** (bracket highlights the region shown in the panels on the right). **h–j**, Whole mount *in situ* hybridization of 56-day-old organoids showing epithelial expression of SOX9 (**h**) and restricted 'crypt-like' expression of the stem cell markers LGR5 (**i**) and ASCL2 (**j**). Insets show sense controls for each probe. Scale bars, 20  $\mu$ m.

which ultimately give rise to the stem cell niche in the crypt of Lieberkühn<sup>15</sup>. iPSCs were equally capable of forming intestinal progenitor domains (Supplementary Fig. 9e). Thus, PSC-derived intestinal epithelium continued to mature *in vitro* and develop proliferative domains with nascent intestinal stem cells.

Between 18 and 28 days in culture, we observed cytodifferentiation of the stratified epithelium into a columnar epithelium containing brush borders and all of the major cell lineages of the gut as determined by immunofluorescence and quantitative polymerase chain reaction with reverse transcription (RT-qPCR) (Fig. 4a–d and Supplementary Fig. 11). By 28 days of culture, villin (Fig. 4a) and DPPIV (not shown) were localized to the apical surface of the polarized columnar epithelium and transmission electron microscopy revealed a brush border of apical microvilli indistinguishable from those found in mature intestine (Fig. 4d and Supplementary Fig. 1). Enterocytes had a functional peptide transport system and were able to absorb a fluorescently



**Figure 4 | Formation and function of intestinal cell types and regulation of enteroendocrine differentiation by NEUROG3.** **a–c**, Twenty-eight-day iPSC-derived organoids were analysed for villin (VIL) and the goblet cell marker mucin (MUC2) (**a**), the Paneth cell marker lysozyme (LYZ) (**b**), or the endocrine cell marker chromogranin A (CHGA) (**c**). Nuc, nuclei. **d**, Electron micrograph showing an enterocyte cell with a characteristic brush border with microvilli (inset). **e**, Epithelial uptake of the fluorescently labelled dipeptide d-Ala-Lys-AMCA (arrowheads) indicating a functional peptide transport system. **f–h**, Adenoviral expression of NEUROG3 (Ad-NEUROG3) causes a fivefold increase in CGA<sup>+</sup> cells compared to a GFP control (Ad-GFP).  $n = 4$  biological samples; \* $P = 0.005$ . **i–k**, Organoids were generated from human ES cells that were stably transduced with shRNA-expressing lentiviral vectors. Compared to control shRNA organoids, NEUROG3 shRNA organoids had a 95% reduction in the number of CHGA<sup>+</sup> cells.  $n = 3$  for shRNA controls and  $n = 5$  for NEUROG3-shRNA; \* $P = 0.018$ . Scale bar in **a** is 10  $\mu$ m; all others are 20  $\mu$ m. Error bars are s.e.m.

labelled dipeptide (Fig. 4e)<sup>22</sup>. Cell counting revealed that the epithelium contained approximately 15% MUC2<sup>+</sup> goblet cells, which secrete mucins into the lumen of the organoid, 18% lysozyme-positive cells, which are indicative of Paneth cells, and ~1% chromogranin-A-expressing enteroendocrine cells (Fig. 4 and Supplementary Fig. 11g). MUC2 and lysozyme staining indicated that the goblet and Paneth cells in 28-day organoids are immature (Fig. 4a, b). However, in organoids that were passaged over 100 days, all cells had acquired a more mature phenotype and Paneth cells were often localized in crypt-like structures (Supplementary Fig. 12b, c). RT-qPCR confirmed the presence of additional markers of differentiated enterocytes (*FABP2*; also known as *IFABP*) and Paneth cells (*MMP7*) (Supplementary Fig. 11). Individual organoids seemed to be a mix of proximal intestine (GATA4<sup>+</sup>/GATA6<sup>+</sup>) and distal intestine (GATA4<sup>+</sup>/GATA6<sup>+</sup>; *HOXA13*-expressing) (Supplementary Figs 11 and 13)<sup>23</sup>. Thus, directed differentiation of PSCs into intestinal tissue *in vitro* is highly efficient in generating three-dimensional intestinal tissue containing crypt-like progenitor niches, villus-like domains and all of the differentiated cell types of the intestinal epithelium.

Intestinal organoids contained a mesenchymal layer that developed along with the epithelium in a staged manner similar to embryonic development<sup>10,24</sup> (Supplementary Fig. 14). Mesenchyme probably came from the 2% of mesoderm cells that were present after activin differentiation, which expanded up to 10% in FGF4-treated hindgut cultures (Supplementary Fig. 2). At 14 days, organoids broadly expressed mesenchymal markers including FOXF1 and vimentin



(Supplementary Fig. 14), similar to an e12.5 embryonic intestine (Supplementary Fig. 7). We also observed vimentin/smooth muscle actin (SMA; also known as ACTA2) double-positive cells indicative of intestinal subepithelial myofibroblasts<sup>25</sup>. By 28 days, we observed a layer of SMA<sup>+</sup>/desmin<sup>+</sup> double-positive cells, indicating smooth muscle, and desmin<sup>+</sup>/vimentin<sup>+</sup> fibroblasts<sup>26</sup>. The fact that intestinal mesenchyme differentiation coincided with differentiation of the overlying epithelium indicates that epithelial–mesenchymal crosstalk may be important in the development of PSC-derived intestinal organoids.

The molecular basis of congenital malformations in humans is often inferred from functional studies in model organisms. For example, neurogenin 3 (*NEUROG3*) was investigated as a candidate gene responsible for congenital loss of intestinal enteroendocrine cells in humans<sup>18</sup> because of its known role in enteroendocrine cell development in mouse<sup>27–30</sup>. However, it has been impossible to directly investigate the role of *NEUROG3* during human intestinal development. We therefore performed gain- and loss-of-function analyses to investigate the role of *NEUROG3* during human enteroendocrine cell development (Fig. 4 and Supplementary Fig. 15). *NEUROG3* was overexpressed in 28-day human organoids using adenoviral (Ad)-mediated transduction<sup>31</sup>. After 7 days, approximately 5% of cells were GFP<sup>+</sup> and Ad-*NEUROG3*–GFP-infected organoids contained fivefold more chromogranin A<sup>+</sup> endocrine cells than control organoids (Ad-enhanced GFP (eGFP)) (Fig. 4f–h and Supplementary Fig. 15), demonstrating that *NEUROG3* expression is sufficient to promote an enteroendocrine cell fate. To knockdown endogenous *NEUROG3*, we generated human ES cell lines by transducing cells with *NEUROG3* short hairpin (sh)RNA-expressing lentiviral vectors. *NEUROG3* mRNA levels were knocked down by 63% and this resulted in a 90% reduction in the number of enteroendocrine cells (Fig. 4i–k and Supplementary Fig. 15d–f), demonstrating that intestinal enteroendocrine cell development is highly dependent on *NEUROG3* expression. This indicates that partial loss-of-function mutations in human *NEUROG3* would be sufficient to cause a marked reduction in enteroendocrine cell numbers.

This is the first report, to our knowledge, demonstrating that human PSCs can be efficiently directed to differentiate *in vitro* into human tissue with a three-dimensional architecture and cellular composition remarkably similar to the fetal intestine. Moreover, PSC-derived human intestinal tissue undergoes maturation *in vitro*, developing intestinal stem cells and acquiring both absorptive and secretory functionality. This system allows for functional studies to investigate the molecular basis of human congenital gut defects *in vitro* and to generate intestinal tissue for eventual transplantation-based therapy for diseases such as necrotizing enterocolitis, inflammatory bowel diseases and short-gut syndromes. The ability to generate human intestinal tissues should also greatly facilitate future studies of intestinal stem cells and drug design to enhance absorption and bioavailability.

## METHODS SUMMARY

**Generation of human intestinal organoids.** Human ES cells and iPSCs were maintained on Matrigel (BD Biosciences) in mTesR1 medium without feeders. Differentiation into definitive endoderm was carried out as previously described<sup>14</sup>. Briefly, a 3-day activin A (R&D systems) differentiation protocol was used. Cells were treated with activin A (100 ng ml<sup>−1</sup>) for three consecutive days in RPMI 1640 medium (Invitrogen) with increasing concentrations of 0%, 0.2% and 2% HyClone defined fetal bovine serum (dFBS; Thermo Scientific). For hindgut differentiation, definitive endoderm cells were incubated in 2% dFBS–DMEM/F12 with 500 ng ml<sup>−1</sup> FGF4 and 500 ng ml<sup>−1</sup> WNT3A (R&D Systems) for up to 4 days. Between 2 and 4 days of treatment with growth factors, three-dimensional floating spheroids formed and were then transferred into three-dimensional cultures previously shown to promote intestinal growth and differentiation<sup>15,16</sup>. Briefly, spheroids were embedded in Matrigel (BD Bioscience) containing 500 ng ml<sup>−1</sup> R-Spondin1 (R&D Systems), 100 ng ml<sup>−1</sup> Noggin (R&D Systems) and 50 ng ml<sup>−1</sup> EGF (R&D Systems). After the Matrigel solidified, medium (advanced DMEM/F12; Invitrogen) supplemented with L-glutamine, 10 μM HEPES, N2 supplement (R&D Systems), B27 supplement (Invitrogen), and penicillin/streptomycin-containing growth factors was overlaid and replaced every 4 days.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 22 April; accepted 23 November 2010.**

**Published online 12 December 2010.**

- Mayhew, C. N. & Wells, J. M. Converting human pluripotent stem cells into  $\beta$ -cells: recent advances and future challenges. *Curr. Opin. Organ Transplant* **15**, 54–60 (2010).
- Spence, J. R. & Wells, J. M. Translational embryology: using embryonic principles to generate pancreatic endocrine cells from embryonic stem cells. *Dev. Dyn.* **236**, 3218–3227 (2007).
- Cai, J. *et al.* Directed differentiation of human embryonic stem cells into functional hepatic cells. *Hepatology* **45**, 1229–1239 (2007).
- D'Amour, K. A. *et al.* Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature Biotechnol.* **24**, 1392–1401 (2006).
- Song, Z. *et al.* Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. *Cell Res.* **19**, 1233–1242 (2009).
- Zhang, D. *et al.* Highly efficient differentiation of human ES cells and iPSCs into mature pancreatic insulin-producing cells. *Cell Res.* **19**, 429–438 (2009).
- Basma, H. *et al.* Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. *Gastroenterology* **136**, 990–999 (2008).
- Touboul, T. *et al.* Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51**, 1754–1765 (2010).
- Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells *in vivo*. *Nature Biotechnol.* **26**, 443–452 (2008).
- Zorn, A. M. & Wells, J. M. Vertebrate endoderm development and organ formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–251 (2009).
- D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature Biotechnol.* **23**, 1534–1541 (2005).
- Dessimoz, J., Opoka, R., Kordich, J. J., Grapin-Botton, A. & Wells, J. M. FGF signaling is necessary for establishing gut tube domains along the anterior-posterior axis *in vivo*. *Mech. Dev.* **123**, 42–55 (2006).
- McLin, V. A., Rankin, S. A. & Zorn, A. M. Repression of Wnt/ $\beta$ -catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development* **134**, 2207–2217 (2007).
- Wells, J. M. & Melton, D. A. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* **127**, 1563–1572 (2000).
- Gracz, A. D., Ramalingam, S. & Magness, S. T. Sox9 expression marks a subset of CD24-expressing small intestine epithelial stem cells that form organoids *in vitro*. *Am. J. Physiol. Gastrointest. Liver Physiol.* **298**, G590–G600 (2010).
- Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- de Santa Barbara, P., van den Brink, G. R. & Roberts, D. J. Development and differentiation of the intestinal epithelium. *Cell. Mol. Life Sci.* **60**, 1322–1332 (2003).
- Wang, J. *et al.* Mutant neurogenin-3 in congenital malabsorptive diarrhea. *N. Engl. J. Med.* **355**, 270–280 (2006).
- Cao, L. *et al.* Intestinal lineage commitment of embryonic stem cells. *Differentiation* doi:10.1016/j.diff.2010.09.182 (in the press).
- Torihashi, S. *et al.* Gut-like structures from mouse embryonic stem cells as an *in vitro* model for gut organogenesis preserving developmental potential after transplantation. *Stem Cells* **24**, 2618–2626 (2006).
- van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
- Groneberg, D. A., Doring, F., Eynott, P. R., Fischer, A. & Daniel, H. Intestinal peptide transport: ex vivo uptake studies and localization of peptide carrier PEPT1. *Am. J. Physiol. Gastrointest. Liver Physiol.* **281**, G697–G704 (2001).
- Haveri, H. *et al.* Transcription factors GATA-4 and GATA-6 in normal and neoplastic human gastrointestinal mucosa. *BMC Gastroenterol.* **8**, 9 (2008).
- McLin, V. A., Henning, S. J. & Jamrich, M. The role of the visceral mesoderm in the development of the gastrointestinal tract. *Gastroenterology* **136**, 2074–2091 (2009).
- Ormestad, M. *et al.* Foxf1 and Foxf2 control murine gut development by limiting mesenchymal Wnt signaling and promoting extracellular matrix production. *Development* **133**, 833–843 (2006).
- Kosinski, C. *et al.* Indian hedgehog regulates intestinal stem cell fate through epithelial–mesenchymal interactions during development. *Gastroenterology* **139**, 893–903 (2010).
- Jenny, M. *et al.* Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *EMBO J.* **21**, 6338–6347 (2002).
- Lee, C. S., Perreault, N., Brestelli, J. E. & Kaestner, K. H. Neurogenin 3 is essential for the proper specification of gastric enteroendocrine cells and the maintenance of gastric epithelial cell identity. *Genes Dev.* **16**, 1488–1497 (2002).
- Lopez-Diaz, L. *et al.* Intestinal Neurogenin 3 directs differentiation of a bipotential secretory progenitor to endocrine cell rather than goblet cell fate. *Dev. Biol.* **309**, 298–305 (2007).
- Ootani, A. *et al.* Sustained *in vitro* intestinal epithelial culture within a Wnt-dependent stem cell niche. *Nature Med.* **15**, 701–706 (2009).
- Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to  $\beta$ -cells. *Nature* **455**, 627–632 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the laboratory, D. Wiginton and C. Wylie for input. We also thank M. Kofron, T. Stefader and R. Lang for assistance with imaging.

Vectors and antibodies were from D. Melton (Addgene no. 19410, 19413), S. Yamanaka (17217–17220), C. Baum (OCT4, KLF4, SOX4, MYC lenti), and I. Manabe (KLF5 antibody). This work was supported by the Juvenile Diabetes Research Foundation JDRF-2-2003-530 (J.M.W.) and NIH, R01GM072915 (J.M.W.); R01DK080823A1 and S1 (A.M.Z. and J.M.W.); R03 DK084167 and R01 CA142826 (N.F.S.), F32 DK83202-01 and T32 HD07463 (J.R.S.). We also acknowledge core support for viral vectors, microarrays (supported by P30 DK078392), karyotyping and the Pluripotent Stem Cell Facility (supported by U54 RR025216).

**Author Contributions** J.M.W. and J.R.S. conceived the study and experimental design, performed and analysed experiments and co-wrote the manuscript. S.A.R., M.F.K. and J.E.V.

performed experiments. C.N.M., M.F.K., K.T., V.V.K., J.E.V., E.E.H. and S.I.W. provided reagents, conceptual and/or technical support in generating and characterizing iPSC lines and intestinal organoids. N.F.S. and A.M.Z. provided additional conceptual and experimental support and co-funded the project. All authors read and approved the final manuscript.

**Author Information** Data have been deposited at NCBI under accession number GSE25557. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.M.W. ([james.wells@cchmc.org](mailto:james.wells@cchmc.org)).

## METHODS

**Maintenance of PSCs.** Human ES cells and induced pluripotent stem cells were maintained on Matrigel (BD Biosciences) in mTesR1 medium<sup>32,33</sup>. Cells were passaged approximately every 4 days, depending on colony density. To passage PSCs, they were washed with DMEM/F12 medium (no serum) (Invitrogen) and incubated in DMEM/F12 with 1 mg ml<sup>-1</sup> dispase (Invitrogen) until colony edges started to detach from the dish. The dish was then washed 3 times with DMEM/F12 medium. After the final wash, DMEM/F12 was replaced with mTesR1. Colonies were scraped off of the dish with a cell scraper and gently triturated into small clumps and passaged onto fresh Matrigel-coated plates.

**Differentiation of PSCs into definitive endoderm.** Differentiation into definitive endoderm was carried out as previously described<sup>11</sup>. Briefly, a 3-day activin A (R&D systems) differentiation protocol was used. Cells were treated with activin A (100 ng ml<sup>-1</sup>) for three consecutive days in RPMI 1640 media (Invitrogen) with increasing concentrations of 0%, 0.2% and 2% HyClone defined fetal bovine serum (dFBS; Thermo Scientific).

**Differentiation of definitive endoderm in permissive media.** After differentiation into definitive endoderm, cells were incubated in DMEM/F12 plus 2% dFBS with either 0, 50 or 500 ng ml<sup>-1</sup> FGF4 and/or 0, 50 or 500 ng ml<sup>-1</sup> WNT3A (R&D Systems) for 6, 48 or 96 h. Cultures were then grown in permissive medium consisting of DMEM plus 10% FBS for an additional 7 days.

**Directed differentiation into hindgut and intestinal organoids.** After differentiation into definitive endoderm, cells were incubated in 2% dFBS-DMEM/F12 with either 50 or 500 ng ml<sup>-1</sup> FGF4 and/or 50 or 500 ng ml<sup>-1</sup> WNT3A (R&D Systems) for 2–4 days. After 2 days with treatment of growth factors, three-dimensional floating spheroids were present in the culture. Three-dimensional spheroids were transferred into an *in vitro* system to support intestinal growth and differentiation previously described<sup>15,16</sup>. Briefly, spheroids were embedded in Matrigel (BD Bioscience; no. 356237) containing 500 ng ml<sup>-1</sup> R-Spondin1 (R&D Systems), 100 ng ml<sup>-1</sup> Noggin (R&D Systems) and 50 ng ml<sup>-1</sup> EGF (R&D Systems). After the Matrigel solidified, medium (advanced DMEM/F12; Invitrogen) supplemented with L-glutamine, 10 µM HEPES, N2 supplement (R&D Systems), B27 supplement (Invitrogen), and penicillin/streptomycin-containing growth factors was overlaid and replaced every 4 days.

**Generation and characterization of iPSC lines.** Normal human skin keratinocytes (HSKs) were obtained from donors with informed consent (Cincinnati Children's Hospital Medical Center (CCHMC) Institutional Review Board protocol CR1\_2008-0899). Normal HSKs were isolated from punch biopsies following trypsinization and subsequent culture on irradiated NIH3T3 feeder cells in F medium<sup>34</sup>. For iPSC generation, normal HSKs were transduced on two consecutive days with a 1:1:1:1 mix of recombinant RD114-pseudotyped retroviruses expressing OCT4, SOX2, KLF4 and MYC<sup>35,36</sup> in the presence of 8 µg ml<sup>-1</sup> polybrene. Twenty-four hours after the second transduction the virus mix was replaced with fresh F medium and cells were incubated for an additional three days. Cells were then trypsinized and seeded into 6-well dishes containing 1.875 × 10<sup>5</sup> irradiated mouse fibroblasts per well and Epilife medium. On the following day, medium was replaced with DMEM/F12 50:50 medium supplemented with 20% knockout serum replacement, 1 mM L-glutamine, 0.1 mM β-mercaptoethanol, 1 × non-essential amino acids, 4 ng ml<sup>-1</sup> basic fibroblast growth factor, and 0.5 mM valproic acid. Morphologically identifiable iPSC colonies arose after 2–3 weeks and were picked manually, expanded and analysed for expression of human PSC markers NANOG, DNMT3B, and using the antigen antibodies Tra1-60 and Tra1-81<sup>37,38</sup>. Early passage iPSC lines were adapted to feeder-free culture conditions consisting of maintenance in mTesR1 (Stem Cell Technologies) in culture dishes coated with Matrigel (BD Biosciences) and lines were karyotyped.

**Microarray analysis of human ES cells, iPSCs and definitive endoderm cultures.** For microarray analysis, RNA was isolated from undifferentiated and 3-day activin-treated human ES cell and iPSC cultures and used to create target DNA for hybridization to Affymetrix Human 1.0 Gene ST Arrays using standard procedures (Affymetrix). Independent biological triplicates were performed for each cell line and condition. Affymetrix microarray Cel files were subjected to RMA normalization in GeneSpring 10.1. Probe sets were first filtered for those that are overexpressed or underexpressed and then subjected to statistical analysis for differential expression by 2 fold or more between undifferentiated and differentiated cultures with *P* < 0.05 using the Students *t*-test. Log2 gene expression ratios

were then subjected to hierarchical clustering using the standard correlation distance metric as implemented in GeneSpring.

**Adenoviral-mediated expression of NEUROG3.** Adenoviral plasmids were obtained from Addgene and particles were generated as previously described<sup>31</sup>. Transduction was done on 28-day organoids that were removed from Matrigel, manually bisected then incubated in Ad-GFP or Ad-NEUROG3 viral supernatant and medium at a 1:1 ratio for approximately 4 h. Organoids were then re-embedded in Matrigel and incubated overnight with viral supernatant and medium at a 1:1 ratio. The next day, fresh organoid medium was placed on the cultures and was changed as described until the end of the experiment.

**shRNA knockdown human ES cell lines.** GpZ shRNA lentiviral vectors were obtained from Open Biosystems (GpZ-NEUROG3 Open Biosystems clone no. v2lhs\_309089; v2lhs\_309091; v2lhs\_309093; v2lhs\_309092 and GpZ-Control; Openbiosystems clone no. RHS4346). The CCHMC Viral Vector Core produced high-titre lentiviral particles for each plasmid. Low-passage H9 human ES cells were dissociated into a single-cell suspension using Accutase, were spun down and resuspended in mTesR1 containing 10 µM Y-27632. Cells were plated at low density and incubated with lentivirus for 24 h. For the NEUROG3 shRNA knockdown line, particles from all four vectors were used. mTesR1 was replaced daily, and after 72 h selection for puromycin- (2–4 µg ml<sup>-1</sup>) resistant human ES cells was carried out. Puromycin-resistant colonies were routinely maintained and passaged in mTesR1 + puromycin (4 µg ml<sup>-1</sup>).

**β-Ala-Lys-AMCA uptake.** β-Ala-Lys-AMCA was purchased from BioTrend Chemicals and was resuspended in water. Intestinal organoids were cut in half using a scalpel and were incubated for four hours in advanced DMEM/F12 plus 24 µM β-Ala-Lys-AMCA. Following incubation, organoids were washed several times in PBS, embedded in OCT freezing medium and were frozen at –70 °C. Ten-micrometre cryosections were cut and processed for standard immunohistochemistry.

**Tissue processing, immunohistochemistry and microscopy.** Tissues were fixed for 1 h to overnight in 4% paraformaldehyde or 3% glutaraldehyde for transmission electron microscopy (TEM). Cultured PSCs and definitive endoderm cells were stained directly. Hindgut and intestinal organoids were embedded in paraffin, epoxy resin LX-112 (Ladd Research), or frozen in OCT. Sections were cut at 6–10 µm for standard microscopy and 0.1 µm for TEM. TEM sections were stained with uranyl acetate. Paraffin sections were deparaffinized, subjected to antigen retrieval, blocked in the appropriate serum (5% serum in 1 × PBS plus 0.5% Triton-X) for 30 min, and incubated with primary antibody overnight at 4 °C. Slides were washed and incubated in secondary antibody in blocking buffer for 2 h at room temperature (23 °C). For a list of antibodies used and dilutions, see Supplementary Table 3. Slides were washed and mounted using Fluoromount-G. Confocal images were captured on a Zeiss LSM510 and Z-stacks were analysed and assembled using AxioVision software. An Hitachi H7600 transmission electron microscope was used to capture images.

**RNA isolation, RT-qPCR.** RNA was isolated using the Nucleospin II RNA isolation kit (Clontech). Reverse transcription was carried out using the SuperScriptIII Supremix (Invitrogen) according to manufacturer's protocol. Finally, qPCR was carried out using Quantitect SybrGreen MasterMix (Qiagen) on a Chromo4 Real-Time PCR (BioRad). PCR primers sequences were typically obtained from qPrimerDepot (<http://primerdepot.nci.nih.gov/>). Primer sequences are available upon request.

32. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nature Methods* **3**, 637–646 (2006).
33. Ludwig, T. E. *et al.* Derivation of human embryonic stem cells in defined conditions. *Nature Biotechnol.* **24**, 185–187 (2006).
34. Lambert, P. F. *et al.* Using an immortalized cell line to study the HPV life cycle in organotypic "raft" cultures. *Methods Mol. Med.* **119**, 141–155 (2005).
35. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
36. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
37. Richards, M., Tan, S. P., Tan, J. H., Chan, W. K. & Bongso, A. The transcriptome profile of human embryonic stem cells as defined by SAGE. *Stem Cells* **22**, 51–64 (2004).
38. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).



# Timing of plant immune responses by a central circadian regulator

Wei Wang<sup>1\*</sup>, Jinyoung Yang Barnaby<sup>1\*†</sup>, Yasuomi Tada<sup>1†</sup>, Hairi Li<sup>2</sup>, Mahmut Tör<sup>3</sup>, Daniela Caldelari<sup>1†</sup>, Dae-un Lee<sup>1</sup>, Xiang-Dong Fu<sup>2</sup> & Xinnian Dong<sup>1</sup>

**The principal immune mechanism against biotrophic pathogens in plants is the resistance (R)-gene-mediated defence<sup>1</sup>. It was proposed to share components with the broad-spectrum basal defence machinery<sup>2</sup>. However, the underlying molecular mechanism is largely unknown. Here we report the identification of novel genes involved in R-gene-mediated resistance against downy mildew in *Arabidopsis* and their regulatory control by the circadian regulator, CIRCADIAN CLOCK-ASSOCIATED 1 (CCA1). Numerical clustering based on phenotypes of these gene mutants revealed that programmed cell death (PCD) is the major contributor to resistance. Mutants compromised in the R-gene-mediated PCD were also defective in basal resistance, establishing an interconnection between these two distinct defence mechanisms. Surprisingly, we found that these new defence genes are under circadian control by CCA1, allowing plants to ‘anticipate’ infection at dawn when the pathogen normally disperses the spores and time immune responses according to the perception of different pathogenic signals upon infection. Temporal control of the defence genes by CCA1 differentiates their involvement in basal and R-gene-mediated defence. Our study has revealed a key functional link between the circadian clock and plant immunity.**

The life cycles of biotrophic pathogens of plants are intimately linked with host metabolism controlled by the cycle of day and night. Hence, their interactions with the host may be dictated by the circadian clock. This is especially likely as plants do not have specialized immune cells and their immune responses have to be finely balanced with other cellular functions. However, a link between the circadian clock and plant defence has never been firmly established<sup>3</sup>.

The major plant defence strategy against biotrophic pathogens is resistance (R)-gene-mediated immunity. Detection of a pathogen-encoded virulence effector by the R protein triggers programmed cell death (PCD) and several other physiological responses collectively known as the hypersensitive response<sup>4</sup>. The effector-specific R-gene-mediated resistance may share components with the broad-spectrum basal defence machinery<sup>2</sup>. But how these components are differentially regulated is still unclear.

We chose to study resistance against *Hyaloperonospora arabidopsidis* (*Hpa*) because this obligate biotrophic oomycete pathogen causes downy mildew disease on *Arabidopsis* leaves through clearly defined infection steps<sup>5</sup>, allowing better dissection of the corresponding resistance mechanisms blocking these steps. The *Arabidopsis* Columbia (Col-0) accession is resistant to the *Hpa* Emwa1 isolate due to the presence of the R gene, *RPP4* (ref. 6). We performed a time-course expression profiling of wild type and *rpp4* (Supplementary Fig. 1) in response to *Hpa* Emwa1 infection. Based on the phenotype development and defence marker gene expression (Supplementary Fig. 2a, b), we identified 106 genes differentially expressed in wild type and *rpp4* at

2 days post inoculation (dpi) (Supplementary Fig. 2c). These candidate genes were induced earlier than the previously reported immune regulators including *EDS5*, *PAD4*, *PBS3*, *ICS1*, *NDR1* and *EDS1* (ref. 7), which were known to function downstream of R gene activation.

We inoculated the T-DNA insertion mutants (from the ABRC and NASC Stock Centres) of these 106 candidate genes with *Hpa* Emwa1 and identified 22 mutants that displayed enhanced susceptibility compared to wild type based on sporangiophore growth and other disease symptoms (for example, chlorosis) by microscopic inspection. For most of the 22 genes, at least two homozygous mutant T-DNA alleles were tested (Supplementary Fig. 3 and Supplementary Tables 1 and 2).

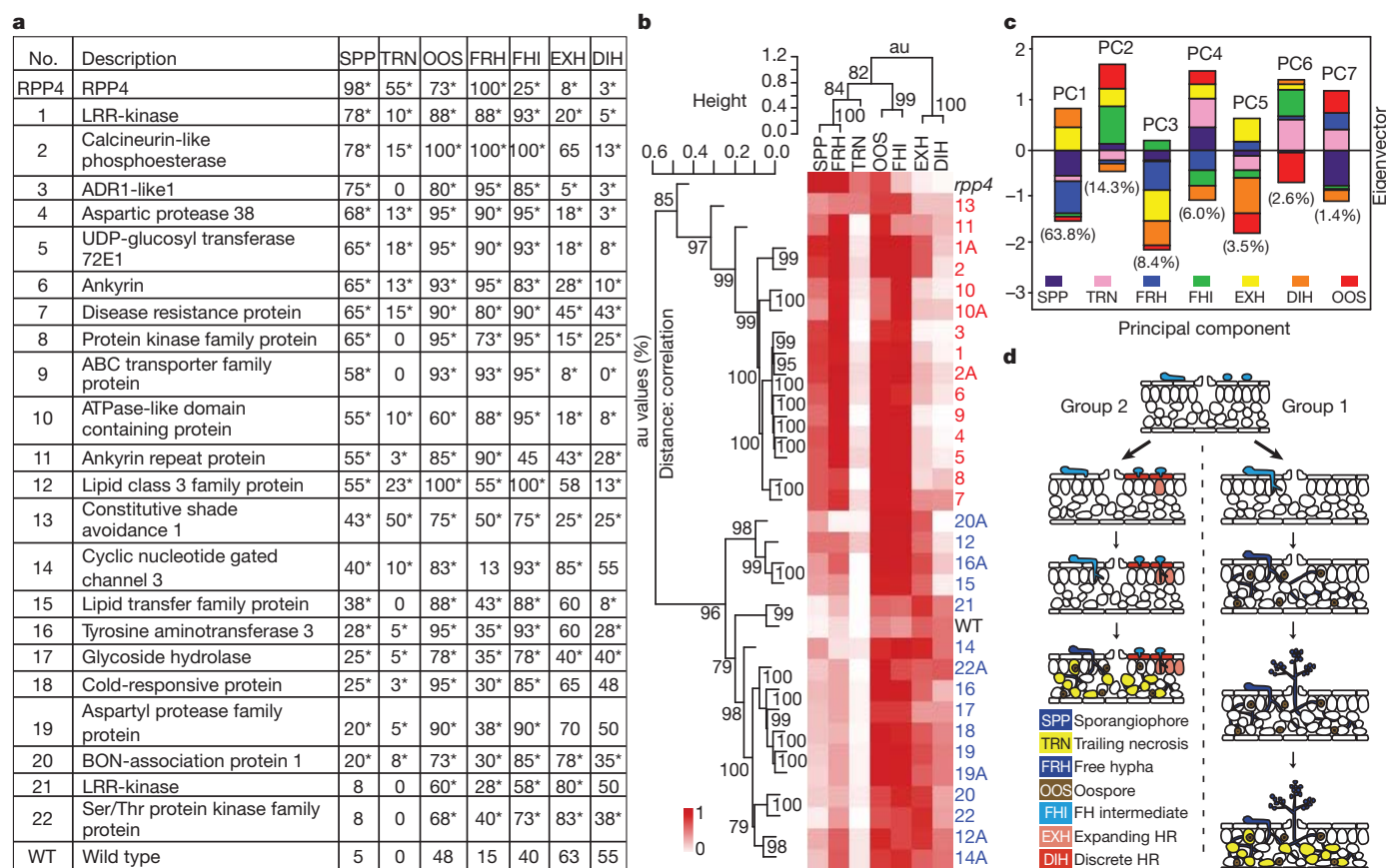
To identify specific resistance defects in the mutants, we stained the infected plants with lactophenol trypan blue (LTB) 7 dpi and scored for the occurrence of the seven phenotypes represented in Supplementary Fig. 4a. As shown in Supplementary Fig. 4b, the *rpp4* mutant had the highest percentage of leaves with sporangiophores (SPP), confirming that its resistance to *Hpa* Emwa1 is completely compromised as SPP indicates completion of the infection cycle. Wild type had the highest score of discrete hypersensitive response (DIH), which was defined by the small cluster of infected host cells that underwent PCD, a phenotype associated with R-gene-mediated resistance. The phenotype scores are also presented numerically in Fig. 1a and the mutants are ranked on the basis of their SPP scores.

Hierarchical clustering of the mutants using their phenotype scores (Fig. 1a) put these 22 gene mutants into two groups (Fig. 1b). Similar groupings are also obtained with data from three biological replicates (Supplementary Fig. 5). Eigenvectors derived from principal component analysis indicate that 63.8% of the phenotype variations could be accounted for by PC1 with indicators of resistance, DIH and expanding hypersensitive response (EXH), as positive contributors and disease phenotypes, SPP and free hypha (FRH), as negative contributors (Fig. 1c). If PC2 was also considered, six out of the seven phenotypes had significant contributions.

The Group 1 mutants (red numbers in Fig. 1b and Supplementary Fig. 5) seem to be defective in R-gene-mediated PCD (low DIH and EXH scores) with high disease symptoms (FRH and SPP). In contrast, the Group 2 mutants (blue numbers) appeared to be intact in PCD with high EXH and DIH scores and milder symptoms (low FRH and SPP scores). To determine the resistance defects in Group 2 mutants, we examined them for other R-gene-mediated physiological responses, such as accumulation of phenolic compounds involved in cell wall strengthening against pathogen penetration and deposition of callose after *Hpa* Emwa1 inoculation. We found that Mutant 16 (tyrosine aminotransferase 3) was defective in phenolic compound accumulation at the site of pathogen penetration (Supplementary Fig. 6a) and Mutants 12 (lipase class 3 family protein) and 14 (cyclic nucleotide gated channel 3) showed a deficiency in callose deposition in

<sup>1</sup>Department of Biology, P. O. Box 90338, Duke University, Durham, North Carolina 27708, USA. <sup>2</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA. <sup>3</sup>National Pollen and Aerobiology Research Unit (NPAP), University of Worcester, Henwick Grove, Worcester WR2 6AJ, UK. <sup>†</sup>Present addresses: Crop Systems and Global Change Laboratory, United States Department of Agriculture, Agricultural Research Service, Plant Sciences Institute, Room 342, Building 001, BARC-West, 10300 Baltimore Avenue, Beltsville, Maryland 20705, USA (J.Y.B.); Life Science Research Centre, Institute of Research Promotion, Kagawa University, 2393 Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0795, Japan (Y.T.); Plant Molecular Biology, University of Lausanne, CH-1015 Lausanne, Switzerland (D.C.).

\*These authors contributed equally to this work.



**Figure 1 | Phenotypic analyses discovered two distinct RPP4-mediated resistance responses against *Hpa* Emw1. a**, Phenotype scores (percentage in 40 leaves per genotype). SPP, sporangioaphore; TRN, trailing necrosis; OOS, oospore; FRH, free hypha; FHI, free hyphal intermediate; EXH, expanding hypersensitive response; DIH, discrete hypersensitive response. \* $P < 0.05$ . **b**, Mutants were clustered on the basis of their phenotype scores in Fig. 1a.

response to *Hpa* Emw1 similar to that observed in *rpp4* (Supplementary Fig. 6b).

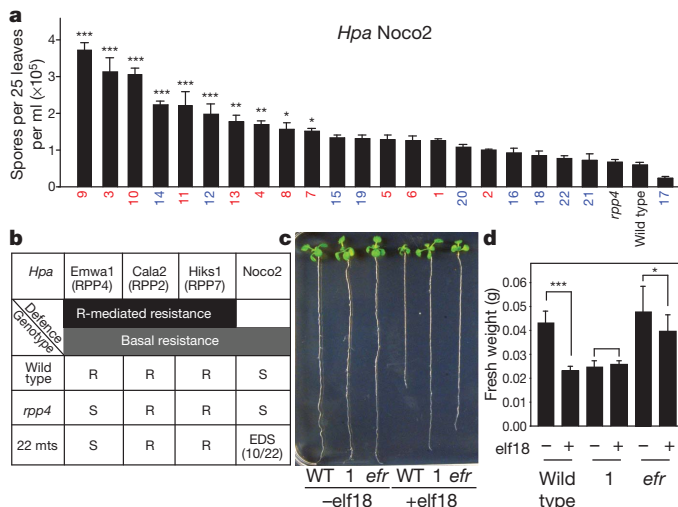
Collectively, these observations suggest that RPP4 regulates at least two separate responses (Fig. 1d): Group 1 genes are required for R-mediated PCD, as mutations in these genes led to low DIH and EXH scores and formation of FRH, TRN, and SPP. The Group 2 genes are probably involved in defence responses other than PCD, such as callose deposition and phenolic compound accumulation. Loss of these latter functions resulted in pathogen penetration even in the presence of PCD. One important conclusion from these data is that PCD is the predominant resistance response against *Hpa* Emw1 because the Group 1 mutants were more susceptible, on the basis of the SPP scores (except Mutant 12), than the Group 2 mutants (Fig. 1a, b). This is supported by the eigenvector composition where PC1 (DIH and EXH) was the major contributor to the phenotypic variations (Fig. 1c). This finding is consistent with the fact that *Hpa* Emw1 is an obligate biotrophic pathogen. Suicidal death of the host cell means the end of the pathogen life cycle. The functional diversity of the Group 1 genes indicates that RPP4-mediated PCD is orchestrated by changes in multiple biological processes, rather than a single triggering event.

We also subjected the 22 defence gene mutants to infection by the virulent isolate, *Hpa* Noco2, to which a cognate *R* gene is absent in Col-0. We found that 10 of the mutants displayed significantly enhanced disease susceptibility (Fig. 2a) demonstrating that these defence genes are involved in both *R* gene-specific PCD and general basal resistance. To determine whether the observed defect in *Hpa* Emw1 resistance is RPP4-specific or due to compromised basal defence, we infected the mutants with *Hpa* isolates, Cala2 and Hiks1, which are known to have

Second allele, 'A'. Group 1, red; Group 2, blue. au, Approximately unbiased *P*-values (0–100%, the higher the number the more significant). **c**, Eigenvectors derived from PCA. The percentage of phenotypic variations captured by each PC is shown. **d**, A diagram showing that the Group 1 mutants are defective in RPP4-mediated PCD, whereas the Group 2 mutants are compromised in formation of physical/chemical barriers with intact PCD.

cognate *R* genes, *RPP2* and *RPP7*, respectively, in the Col-0 background<sup>7,8</sup>. None of the mutants showed compromised resistance (Fig. 2b and Supplementary Table 1) indicating that the deficiency in resistance against *Hpa* Emw1 is RPP4 gene-specific. However, a few of the mutants did show defects in RPS2-mediated resistance to a bacterial pathogen *Pseudomonas syringae* pv. *maculicola* ES4326 carrying AvrRpt2 (*Psm* ES 4326/AvrRpt2) (Supplementary Fig. 7a). Three of them were also hypersusceptible to *Psm* ES4326 in the absence of the AvrRpt2 signal (Supplementary Fig. 7b).

We next subjected all of the 22 mutants to microbial-associated molecular pattern (MAMP) treatments including EF-Tu (elf18) and flagellin (flg22) to examine the interconnection between *R*-gene-mediated resistance and MAMP-triggered basal immunity. We found that Mutant 1, mutated in the leucine-rich repeat receptor-like kinase (LRR-RLK; AT1G35710), was insensitive to elf18 (Fig. 2c, d). Because Mutant 1 also showed the highest level of susceptibility to *Hpa* Emw1 (Fig. 1a), we propose that this LRR-RLK is a link between MAMP-signalling and RPP4-mediated PCD and resistance. Although MAMP-triggered immunity is not typically associated with PCD, MAMP signalling components have been implicated previously in PCD. Mutation of the *Arabidopsis* BRI1-associated receptor kinase 1 (BAK1), which is a MAMP-coreceptor required for responses to elf18 and flg22 (ref. 9), was shown to cause spreading necrosis upon pathogen challenge, indicating that BAK1 is an inhibitor of PCD<sup>10</sup>. However, silencing BAK1 in *Nicotiana benthamiana* blocked the cell death induced by the oomycete elicitor INF1 (ref. 11), consistent with our finding that a MAMP signalling component is involved in PCD resistance against oomycete infection.



**Figure 2 | Some of the RPP4-mediated resistance mutants are also compromised in basal defence.** **a**, Enhanced disease susceptibility to *Hpa* Noco2 based on sporangiospore count 7 dpi ( $n = 3$ ). **b**, Summary of the infection tests on the 22 defence gene mutants (22 mts) using different *Hpa* isolates. S, susceptible; R, resistant; EDS, enhanced disease susceptibility. **c**, Root length measurements 9 days after elf18 treatment ( $n = 3$ ). **d**, Fresh weight measurements 6 days after elf18 treatment ( $n = 3$ ). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

Our genetic data showed that R-mediated resistance and basal defence share common components. This raises the question of how activation of similar sets of genes causes PCD in RPP4-specific resistance against *Hpa* Emwa1 and non-specific basal resistance against *Hpa* Noco2. To understand the differential regulation of these immune mechanisms, we analysed the promoter regions of these 22 genes. Using the Athena program (<http://www.bioinformatics2.wsu.edu/Athena/>), we found significant enrichment of the 'evening element', which is regulated both positively and negatively by the circadian regulator, CCA1 (refs 12, 13). Further examination showed that 14 of the 22 genes contain either evening element and/or the CCA1-binding site and/or have rhythmic expression patterns (Fig. 3a)<sup>14</sup>. Interestingly, the promoter region of *RPP4* also contains two evening elements and its expression shows a circadian rhythm.

To confirm the involvement of the circadian clock in defence, we first examined the responses of clock mutants to *Hpa* Emwa1. The infection was carried out at dawn, the time when *Hpa* spores are normally disseminated in nature<sup>15</sup>. The *cca1* mutant (Salk\_067780) and *ztl-4* (a mutant of *ZEITLUPE*)<sup>16</sup> showed compromised resistance whereas a CCA1-overexpression line (CCA1<sup>OE</sup>)<sup>17</sup> showed enhanced resistance (Fig. 3b). Surprisingly, *lhy*, the mutant of the CCA1 homologue, *LATE AND ELONGATED HYPOCOTYL (LHY)*<sup>18</sup> responded as wild type.

We next examined the expression patterns of all 22 defence genes in wild type, *rpp4* and *cca1* every 2 h in a 46-h time-course, with and without infection by *Hpa* Emwa1. Because of the large number of samples involved, we used the innovative high throughput RNA annealing selection ligation-sequencing (RASL-seq) technology<sup>19</sup> for expression analysis (Supplementary Table 3 and Methods).

As shown in Fig. 3c, consistent with the genetic data, the rhythmic expression of *LHY* was not significantly perturbed by infection in either wild type or *rpp4*. This indicates that RPP4-mediated defence does not disrupt the overall running of the clock, but rather engages CCA1. This specific sensitivity of CCA1 to infection conditions was confirmed using a transgenic line expressing *CCA1:LUC* (ref. 16, Supplementary Fig. 8a). To eliminate the effects of light changes on CCA1 expression, we also performed infection in transgenic plants carrying the *CCA1:LUC* and *LHY:LUC* reporters<sup>16</sup> under the free-running light

cycles (Supplementary Fig. 8b). Similar to the RASL-seq results, *LHY:LUC* expression remained unchanged, whereas *CCA1:LUC* expression was significantly induced and became arrhythmic upon *Hpa* Emwa1 challenge.

Conveniently, the stable expression pattern of *LHY* served as an internal control for the quality of RNA preparations and RASL-seq. Based on the non-negative matrix factorization (NMF) algorithm<sup>20</sup>, the 22 RPP4-regulated genes fit best into two clusters (Fig. 3a and Supplementary Figs 9 and 10). The membership distance of each gene to its cluster is illustrated by the circle radius in Supplementary Fig. 11a. These two clusters corresponded roughly to the two phenotypic groups determined through genetic analysis (Figs 1 and 3a). Most of the Cluster 1 genes containing evening element in their promoters are involved in R-gene-mediated PCD and were therefore the focus of further concern (Fig. 3c). The expression patterns of the Cluster 2 genes are shown in Supplementary Fig. 11b.

Consistent with the fact that evening element is enriched in the Cluster 1 gene promoters (Fig. 3a), the weighted mean expression of these genes largely overlaps with the expression patterns of *CCA1* (Fig. 3c). In wild-type control (Col CK), Cluster 1 genes showed a rhythmic expression pattern with a single sharp peak every evening. In *cca1* (cca1 CK), the expression peaks were greatly diminished, confirming that CCA1 is an activator of these defence genes.

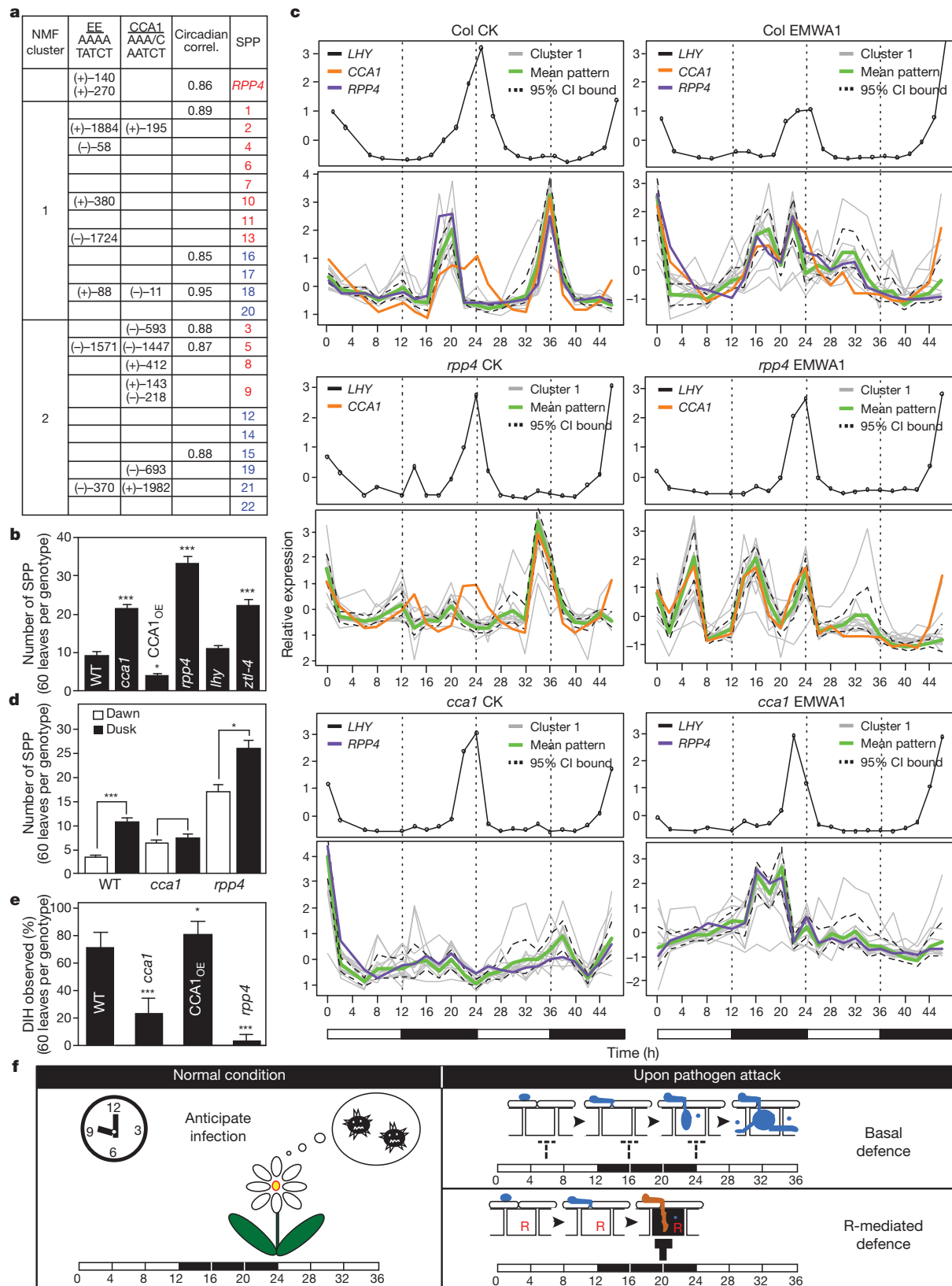
The rhythmic expression of the defence genes in the absence of pathogen indicates that plants are programmed to 'anticipate' infection according to a circadian schedule. The CCA1-mediated pulse expression of the defence genes coincides with the time of *Hpa* sporulation which mainly occurs at night and the time of spore dissemination which takes place at dawn<sup>15</sup>. To test this, we performed *Hpa* Emwa1 infection not only at the normal 'dawn' infection time but also at 'dusk'. We found that if the plants were inoculated at dusk, when infection was unexpected, significantly higher levels of susceptibility were observed in both wild type and *rpp4* (Fig. 3d). CCA1 clearly has a role in conferring resistance at dawn because in *cca1*, more *Hpa* Emwa1 growth was observed compared to wild type. However, no further increase in susceptibility was observed in *cca1* if inoculation was carried at dusk because CCA1 and the CCA1-regulated defence genes are not expressed at this time.

In response to *Hpa* Emwa1 infection, Cluster 1 genes showed drastically different expression patterns in wild type (Col EMWA1) and *rpp4* (*rpp4* EMWA1) (Fig. 3c). Without RPP4, the expression of the defence genes peaked at the 6-, 16- and 24-h time points which coincided with the expected time of *Hpa* spore germination, formation of penetration hyphae and establishment of primary haustoria in mesophyll cells, respectively<sup>5</sup>. This pattern of expression may explain how these defence genes contribute to the basal resistance against *Hpa* infection. Consistent with CCA1 having a role in basal resistance, CCA1<sup>OE</sup> was more resistant to *Hpa* Noco2 than wild type (Supplementary Fig. 12). However, understanding the signalling events between the pathogen and CCA1 leading to this specific timing of defence gene expression will require future research.

In the presence of RPP4, the 6-h expression peak was diminished (Col EMWA1, Fig. 3c). The subsequent perception of the pathogen effector by RPP4 led to the gradual and sustained expression of defence genes. We propose that the prolonged expression of these defence genes, which are normally pulse-expressed at dawn, results in PCD of the infected host cells and pathogen resistance. Consistent with it being a key positive regulator of Cluster 1 defence genes involved in RPP4-mediated PCD, knocking out CCA1 function significantly lowered the average DIH score (a measure of host cell death) in *cca1* after *Hpa* Emwa1 infection (Fig. 3e).

How RPP4 interacts with CCA1 to control the defence gene expression requires further investigation. The spatial resolution of the expression data from the *Hpa* Emwa1-infected samples, homogenized from both infected and uninfected cells, was not enough to allow detailed dissection of the contribution of RPP4, CCA1 and unknown pathogenic





**Figure 3 | The circadian regulator, CCA1, controls the defence gene expression and the timing of immune responses.** **a**, Enrichment of evening element (EE) ( $P < 10^{-5}$ ). NMF, non-negative matrix factorization; CCA1, CCA1-binding sites; Circadian correl., circadian correlations<sup>14</sup>. +, sense; -, antisense. **b**, SPP count 7 dpi by *Hpa* Emwa1 ( $n = 3$ ). **c**, Time-course expression of NMF Cluster 1 genes. CI, confidence interval; CK, control; EMWA1, *Hpa*

Emwa1 inoculated. White bars, day; black bars, night. **d**, SPP count after *Hpa* Emwa1 infection at dawn or dusk ( $n = 3$ ). **e**, Occurrence of DIH 7 dpi by *Hpa* Emwa1. **f**, A model showing circadian regulation of the defence genes in anticipation of infection under normal conditions, in basal and R-gene-mediated resistance. The blocked arrows represent defence against infection.

signals. Nevertheless, RPP4 clearly is not only a target gene of CCA1, but also a partner of CCA1 in regulating the defence genes as their patterns of expression were disturbed in *rpp4* (Fig. 3c and Supplementary Fig. 2c).

Establishment of a molecular link between the plant circadian clock and R-mediated defence reveals a new interface between the plant host and biotrophic pathogens. Although the interactions between R genes and the circadian clock have yet to be studied genetically and at the molecular level, this study indicates a central role of the circadian clock in balancing growth and defence. As summarized in Fig. 3f, we hypothesize that the Cluster 1 genes are pulse-expressed to minimize adverse effects to the host in anticipation of infection under normal conditions and during basal defence. In contrast, detection of a pathogenic effector by the R protein may disrupt this control, leading to PCD of the infected cell and R-mediated resistance, which is a much stronger and signal-specific immune response. There is also an increasing body of evidence indicating that animal immune response is influenced by the circadian clock<sup>21</sup>. Understanding the molecular link between the circadian clock and immunity therefore has broad implications in biology.

## METHODS SUMMARY

*Hyaloperonospora arabidopsidis* (*Hpa*) propagation and inoculation were performed as described<sup>6,22</sup>. Ten-day-old plants were inoculated with the asexual spores suspension ( $5 \times 10^5$  spores per ml) of *Hpa*. Unless specified, the *Hpa* infection was always performed at dawn of the growth chamber's photoperiod. *Hpa* Emwa1-inoculated samples were collected at 0, 0.5, 2 and 4 days post inoculation (dpi). ATH1 GeneChip (Affymetrix) was used for microarray. The arrays were normalized and analysed as described previously<sup>23</sup>. Disease phenotypes were scored after trypan blue staining at 7 dpi<sup>24</sup>. Significance of the phenotypic scores was determined based on binomial distribution. Disease phenotypic analysis was performed using hierarchical clustering with distance measured by the standard correlation (average linkage; scale 0–1). The significance of the clustering (bootstrap 100,000 times) was measured by the approximately unbiased *P*-values (0–100%, the higher the number the more significant<sup>25</sup>). Callose deposition was detected after aniline blue staining<sup>26</sup>. Accumulation of phenolic compounds was examined under ultraviolet illumination (Leica). Root length and fresh weight assays for *elf18* sensitivity were performed as described previously<sup>9</sup>. The evening element enrichment was determined based on hypergeometric distribution. Samples for RASL-seq were prepared according to ref. 19. Non-negative matrix factorization algorithm was used to cluster the genes<sup>20</sup>. RNA extraction was performed as described previously<sup>27</sup>. cDNA synthesis (SuperScript III, Invitrogen) and quantitative PCR (SYBR Green, Qiagen) were performed according to the manufacturer's protocols. For *Pseudomonas* infection, 4-week-old plants were inoculated with 10 mM MgCl<sub>2</sub> or *Pseudomonas syringae* *maculicola* ES4326 with or without the effector *AvrRpt2* (OD<sub>600</sub> = 0.001). The *in planta* bacterial growth was measured at 3 dpi. For diurnal luciferase measurement, protein was extracted and bioluminescence intensity was measured using the Luciferase Assay System (Promega) according to manufacturer's protocol. Ten-day-old plate-grown plants were used for free-running test (details in Methods).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 3 June; accepted 17 December 2010.

1. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
2. Tao, Y. *et al.* Quantitative nature of *Arabidopsis* responses during compatible and incompatible interactions with the bacterial pathogen *Pseudomonas syringae*. *Plant Cell* **15**, 317–330 (2003).
3. Roden, L. C. & Ingle, R. A. Lights, rhythms, infection: the role of light and the circadian clock in determining the outcome of plant–pathogen interactions. *Plant Cell* **21**, 2546–2552 (2009).
4. Lam, E., Kato, N. & Lawton, M. Programmed cell death, mitochondria and the plant hypersensitive response. *Nature* **411**, 848–853 (2001).
5. Donofrio, N. M. & Delaney, T. P. Abnormal callose response phenotype and hypersusceptibility to *Peronospora parasitica* in defence-compromised *Arabidopsis* *nim1-1* and salicylate hydroxylase-expressing plants. *Mol. Plant Microbe Interact.* **14**, 439–450 (2001).

6. Holub, E. B., Beynon, J. L. & Crute, I. R. Phenotypic and genotypic characterization of interactions between isolates of *Peronospora parasitica* and accessions of *Arabidopsis thaliana*. *Mol. Plant Microbe Interact.* **7**, 223–239 (1994).
7. van der Biezen, E. A., Freddie, C. T., Kahn, K., Parker, J. E. & Jones, J. D. *Arabidopsis* RPP4 is a member of the RPP5 multigene family of TIR-NB-LRR genes and confers downy mildew resistance through multiple signalling components. *Plant J.* **29**, 439–451 (2002).
8. McDowell, J. M. *et al.* Downy mildew (*Peronospora parasitica*) resistance genes in *Arabidopsis* vary in functional requirements for NDR1, EDS1, NPR1 and salicylic acid accumulation. *Plant J.* **22**, 523–529 (2000).
9. Chinchilla, D. *et al.* A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature* **448**, 497–500 (2007).
10. Kemmerling, B. *et al.* The BRI1-associated kinase 1, BAK1, has a brassinolide-independent role in plant cell-death control. *Curr. Biol.* **17**, 1116–1122 (2007).
11. Heese, A. *et al.* The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proc. Natl Acad. Sci. USA* **104**, 12217–12222 (2007).
12. Harmer, S. L. *et al.* Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**, 2110–2113 (2000).
13. Harmer, S. L. & Kay, S. A. Positive and negative factors confer phase-specific circadian regulation of transcription in *Arabidopsis*. *Plant Cell* **17**, 1926–1940 (2005).
14. Michael, T. P. *et al.* Network discovery pipeline elucidates conserved time-of-day-specific cis-regulatory modules. *PLoS Genet.* **4**, e14 (2008).
15. Slusarenko, A. & Schlaich, N. L. Downy Mildew of *Arabidopsis thaliana* caused by *Hyaloperonospora parasitica* (formerly *Peronospora parasitica*). *Mol. Plant Pathol.* **4**, 159–170 (2003).
16. Salome, P. A. & McClung, C. R. PSEUDO-RESPONSE REGULATOR 7 and 9 are partially redundant genes essential for the temperature responsiveness of the *Arabidopsis* circadian clock. *Plant Cell* **17**, 791–803 (2005).
17. Wang, Z. Y. & Tobin, E. M. Constitutive expression of the CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) gene disrupts circadian rhythms and suppresses its own expression. *Cell* **93**, 1207–1217 (1998).
18. Schaffer, R. *et al.* The late elongated hypocotyl mutation of *Arabidopsis* disrupts circadian rhythms and the photoperiodic control of flowering. *Cell* **93**, 1219–1229 (1998).
19. Yeakley, J. M. *et al.* Profiling alternative splicing on fiber-optic arrays. *Nature Biotechnol.* **20**, 353–358 (2002).
20. Tamayo, P. *et al.* Metagene projection for cross-platform, cross-species characterization of global transcriptional states. *Proc. Natl Acad. Sci. USA* **104**, 5959–5964 (2007).
21. Bryant, P. A., Trinder, J. & Curtis, N. Sick and tired: does sleep have a vital role in the immune system? *Nature Rev. Immunol.* **4**, 457–467 (2004).
22. Tor, M. *et al.* *Arabidopsis* SGT1b is required for defense signaling conferred by several downy mildew resistance genes. *Plant Cell* **14**, 993–1003 (2002).
23. Brady, S. M. *et al.* A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**, 801–806 (2007).
24. Bowling, S. A., Clarke, J. D., Liu, Y., Klessig, D. F. & Dong, X. The *cpr5* mutant of *Arabidopsis* expresses both NPR1-dependent and NPR1-independent resistance. *Plant Cell* **9**, 1573–1584 (1997).
25. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
26. Adam, L. & Somerville, S. C. Genetic characterization of five powdery mildew resistance loci in *Arabidopsis thaliana*. *Plant J.* **9**, 341–356 (1996).
27. Cao, H., Bowling, S. A., Gordon, S. & Dong, X. Characterization of an *Arabidopsis* mutant that is nonresponsive to inducers of systemic acquired resistance. *Plant Cell* **6**, 1583–1592 (1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Brady for performing mixed model ANOVA of the data and advice on data analyses; J. Li for sharing the protocol for *elf18* treatment; E. Tobin for providing the CCA1<sub>OE</sub> transgenic line; R. McClung for CCA1:LUC, LHY:LUC, *ztl-4* lines; H. Lu for helpful discussion of the work; F. Ausubel, P. Benfey, S. Brady, J. Siedow and R. Mohan for critiquing the manuscript. This work was supported by a grant from NSF (MCB-0519898) to X.D. and a grant (HG004659) to X.-D.F.

**Author Contributions** J.Y.B., W.W., Y.T., D.C. and D.-u.L. identified new components in R-gene-mediated resistance; J.Y.B., W.W. and Y.T. showed that RPP4 controls two major defence responses by phenotypic clustering analysis; J.Y.B., W.W. and M.T. demonstrated that R-gene-mediated resistance shares common components with basal defence machinery; W.W., J.Y.B., H.L., X.-D.F. and X.D. verified the circadian regulator CCA1 plays a key role in timing the different immune responses; W.W., J.Y.B. and X.D. wrote the manuscript with inputs from all co-authors. All authors discussed the results and commented on the manuscript.

**Author Information** The microarray data presented in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE22274. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to X.D. ([xdong@duke.edu](mailto:xdong@duke.edu)).

## METHODS

### *Arabidopsis* and *Hyaloperonospora arabidopsidis* (*Hpa*) growth conditions.

*Arabidopsis* seedlings were grown for 10 days at 16–18 °C, 12-h day length, 80–100% relative humidity before *Hpa* infection through spray of a spore suspension ( $5 \times 10^5$  spores per ml in distilled H<sub>2</sub>O) at dawn according to the photoperiod of the plant growth chamber. *Hpa* Emwa1 and *Hpa* Noco2 were subcultured and inocula prepared using methods modified from previous reports<sup>6,22</sup>.

**RNA extraction and quantitative PCR analysis.** RNA extraction was performed as described previously<sup>27</sup>. cDNA synthesis (SuperScript III, Invitrogen) and quantitative PCR (SYBR Green PCR kit, Qiagen) were performed according to the manufacturer's protocols.

**Microarray.** Ten-day-old wild-type and *rpp4* seedlings were inoculated with *Hpa* Emwa1 and samples were collected at 0, 0.5, 2 and 4 days after inoculation. Total RNA was isolated from the frozen material using the Qiagen RNeasy kit. RNA probes were labelled using the GeneChip Eukaryotic Small Sample Target Labelling Assay Version II and hybridized on the Affymetrix ATH1 GeneChip (Santa Clara). Two biological replicates were performed. The data presented in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE22274 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22274>).

**Normalization and mixed-model analysis.** The mixed-model software used to normalize globally all arrays and to identify differentially expressed probe sets was as described previously<sup>23</sup>. Expression indices were used to calculate *P*- and *q*-values for pairwise comparisons of all probe sets across all treatments. *R*<sup>2</sup> values for the CEL files are as follows (*Col\_0d*: 1 vs 2(0.98); *Col\_0.5d*: 1 vs 2(0.97); *Col\_2d*: 1 vs 2(0.98); *Col\_4d*: 1 vs 2(0.96); *rpp4\_0d*, 1 vs 2 (0.98); *rpp4\_0.5d*, 1 vs 2 (0.97); *rpp4\_2d*, 1 vs 2 (0.98); *rpp4\_4d*, 1 vs 2 (0.97)).

**Phenotyping mutants in response to *Hpa* infection.** Seven days after inoculation with *Hpa* Emwa1 infection, phenotypes were scored following lactophenol trypan blue staining<sup>24</sup>. Leaves were vacuum-infiltrated twice in a solution of phenol, lactic acid, glycerol and water (1:1:1:1) plus 2.5 mg ml<sup>-1</sup> trypan blue. The tubes containing the samples were placed in a boiling water bath for 2 min and allowed to cool for overnight. The leaves were destained in the chloral hydrate solution and then treated with 70% glycerol. Whole leaves were analysed and photographed with a MZ8 stereo microscope (Leica) and a PM-C35 camera (Olympus). Detailed examination of *Hpa* structures was conducted with an Olympus BX60F compound microscope and differential interference contrast (DIC) optics. Leaves were stained for callose as described with modifications<sup>26</sup>. To visualize callose, leaves were cleared in a solution of ethanol and acetic acid (3:1), stained with 0.02% aniline blue in 100 mM sodium phosphate buffer (pH 9) for 1 h, and examined with an Axio imager wide field fluorescence microscope (Zeiss). To detect phenolic compounds, leaves were examined under ultraviolet fluorescent illumination (Leica DMRB). To measure *Hpa* Noco2 infection, infected leaves were collected in 1 ml water, and sporangiospores were counted.

**elf18 and flg22 treatment.** Wild-type *Arabidopsis* (Col-0), *efr* and candidate mutant plants were grown on MS+1% sucrose medium (pH 5.7) plate with 1 μM elf18 or MS alone as a control under continuous light for 9 days for root length assay<sup>9</sup>. For fresh weight assay, plants were grown on MS+1% sucrose medium (pH 5.7) plate for 4 days (16/8 light/dark cycle), transferred into water containing 50 nM elf18 or water as a control, and fresh weight was measured 6 days after elf18 treatment. flg22 (10 nM) was infiltrated into 3-week-old plants for the induction of callose deposition.

**Promoter element analysis.** The statistical significance of over-represented transcription factor (TF) binding elements was calculated using a hypergeometric probability model. The following equation was used to provide the *P*-values:

$$P = 1 - \sum_{x=0}^{k-1} \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

*N* is the total number of promoters in the genome, *n* is the number of promoters in the genome containing the specified TF-binding element, *m* is the size of the selected set of promoters, and *x* is the number of promoters with the specified element in the selected set. Because multiple hypotheses were tested in the analysis, the Bonferroni correction was used. The genome-wide occurrences of these elements in the promoters are used as controls.

**RASL-seq.** The growth conditions (12/12 light/dark cycle, 16–18 °C, 80–100% humidity), which were optimized for *Hpa* infection, were different from those used in traditional circadian studies<sup>12,13</sup>. Samples were collected every 2 h after inoculation and the remaining plants were kept to ensure successful pathogen inoculation. Total RNA for each sample (1 μg) was used for RASL-seq. Primer (gene-specific with flanking 5' or 3' universal sequences) annealing to mRNA and ligation were carried out according to ref. 19. Bar-coded primers were then added to each sample to convert the ligated products to individual libraries, which were pooled from all samples and subjected to multiplex sequencing using Solexa GAII (Illumina).

**RASL-seq data analysis.** The readings from RASL-seq were assumed Poisson distribution. Only those samples with mean readings significantly above zero (*Pr*(mean = 0) < 0.01) were considered for further analysis. The reading for each sample was first divided by the corresponding reading of control, ubiquitin 5 (*UBQ5*; AT3G62250), and then standardized. The resulting matrix was used for clustering analysis.

Non-negative matrix factorization (NMF) algorithm<sup>20</sup> was used to cluster the genes. The number of the clusters was determined by comparing the cophenetic correlation coefficient for a range of cluster numbers (from 2 to 22). The cophenetic correlation coefficient is a measurement of how faithfully the result of NMF clustering preserves the pairwise distances between the original data points. As shown in Supplementary Fig. 9, two clusters generated the highest cophenetic correlation coefficient, which means two clusters can reflect the original data more faithfully than more clusters. Divergence was used as the update rule and cost measurement. Minimum of the data was subtracted from the data matrix to ensure that there were no negative numbers in the matrix. Because the NMF algorithm iteratively updates the decomposition of the data matrix, 300 runs with 10,000 iterations/run were performed to reach the convergence (Supplementary Fig. 10). The membership indicators from NMF clustering were used as weights to calculate the weighted mean expression pattern shown in Fig. 3c. The weights were also used to determine the radii of circles in Supplementary Fig. 11a. Smaller radius indicates a higher membership of the gene to the corresponding cluster.

**Bioluminescence detection.** Protein was extracted and bioluminescence intensity measured using the Luciferase Assay System (Promega) according to the manufacturer's manual. A Victor3 (PerkinElmer) multilabel reader was used to detect the bioluminescence. Substrate (100 μl) was added using an automatic injector. After 3 s shaking, 2 s delay, the signal was captured for 20 s. Log<sub>10</sub> transformation was performed to the raw signals to ensure the normal distribution of the data. After subtraction of the blank, the data were normalized according to the total protein concentrations determined by the Bradford method (Bio-Rad). The resulting data were then standardized.

**Free-running test.** Seeds were sterilized in 2% Plant Preservative Mixture (PPM, Plant Cell Technology) in the dark at 4 °C for 4 days before plating on MS plate (3% sucrose, 1.5% agar) and grown in a 12/12 h light/dark growth chamber for 9 days. At the dawn and the dusk of the ninth day, 2.5 mM luciferin in 0.05% Triton-X 100 was sprayed onto the seedlings. At the dawn of the tenth day, the seedlings were treated by distilled H<sub>2</sub>O or *Hpa* Emwa1 before being placed in a constant light chemiluminescence box. The bioluminescence signals were captured by CCD.



# Oncogenically active *MYD88* mutations in human lymphoma

Vu N. Ngo<sup>1,†\*</sup>, Ryan M. Young<sup>1\*</sup>, Roland Schmitz<sup>1\*</sup>, Sameer Jhavar<sup>1\*</sup>, Wenming Xiao<sup>2\*</sup>, Kian-Huat Lim<sup>1\*</sup>, Holger Kohlhammer<sup>1</sup>, Weihong Xu<sup>1</sup>, Yandan Yang<sup>1</sup>, Hong Zhao<sup>1</sup>, Arthur L. Shaffer<sup>1</sup>, Paul Romesser<sup>1,3</sup>, George Wright<sup>4</sup>, John Powell<sup>2</sup>, Andreas Rosenwald<sup>5</sup>, Hans Konrad Muller-Hermelink<sup>5</sup>, German Ott<sup>6</sup>, Randy D. Gascoyne<sup>7</sup>, Joseph M. Connors<sup>7</sup>, Lisa M. Rimsza<sup>8,9</sup>, Elias Campo<sup>10</sup>, Elaine S. Jaffe<sup>11</sup>, Jan Delabie<sup>12</sup>, Erlend B. Smeland<sup>13</sup>, Richard I. Fisher<sup>9,14</sup>, Rita M. Brazier<sup>9,15</sup>, Raymond R. Tubbs<sup>9,16</sup>, J. R. Cook<sup>9,16</sup>, Denny D. Weisenburger<sup>17</sup>, Wing C. Chan<sup>17</sup> & Louis M. Staudt<sup>1</sup>

The activated B-cell-like (ABC) subtype of diffuse large B-cell lymphoma (DLBCL) remains the least curable form of this malignancy despite recent advances in therapy<sup>1</sup>. Constitutive nuclear factor (NF)- $\kappa$ B and JAK kinase signalling promotes malignant cell survival in these lymphomas, but the genetic basis for this signalling is incompletely understood. Here we describe the dependence of ABC DLBCLs on *MYD88*, an adaptor protein that mediates toll and interleukin (IL)-1 receptor signalling<sup>2,3</sup>, and the discovery of highly recurrent oncogenic mutations affecting *MYD88* in ABC DLBCL tumours. RNA interference screening revealed that *MYD88* and the associated kinases *IRAK1* and *IRAK4* are essential for ABC DLBCL survival. High-throughput RNA resequencing uncovered *MYD88* mutations in ABC DLBCL lines. Notably, 29% of ABC DLBCL tumours harboured the same amino acid substitution, L265P, in the *MYD88* Toll/IL-1 receptor (TIR) domain at an evolutionarily invariant residue in its hydrophobic core. This mutation was rare or absent in other DLBCL subtypes and Burkitt's lymphoma, but was observed in 9% of mucosa-associated lymphoid tissue lymphomas. At a lower frequency, additional mutations were observed in the *MYD88* TIR domain, occurring in both the ABC and germinal centre B-cell-like (GCB) DLBCL subtypes. Survival of ABC DLBCL cells bearing the L265P mutation was sustained by the mutant but not the wild-type *MYD88* isoform, demonstrating that L265P is a gain-of-function driver mutation. The L265P mutant promoted cell survival by spontaneously assembling a protein complex containing *IRAK1* and *IRAK4*, leading to *IRAK4* kinase activity, *IRAK1* phosphorylation, NF- $\kappa$ B signalling, JAK kinase activation of *STAT3*, and secretion of IL-6, IL-10 and interferon- $\beta$ . Hence, the *MYD88* signalling pathway is integral to the pathogenesis of ABC DLBCL, supporting the development of inhibitors of *IRAK4* kinase and other components of this pathway for the treatment of tumours bearing oncogenic *MYD88* mutations.

The current molecular taxonomy of DLBCL distinguishes three main subtypes: ABC, GCB and primary mediastinal B-cell lymphoma (PMBL)<sup>4</sup>. Current therapy is least successful in ABC DLBCL, achieving less than a 40% cure rate<sup>1</sup>. The anti-apoptotic NF- $\kappa$ B signalling pathway is constitutively active in ABC DLBCL owing to oncogenic *CARD11* mutations or chronic active B-cell receptor signalling, augmented by inactivation of *A20*<sup>5–8</sup>. A subset of ABC DLBCLs use JAK

kinase signalling to activate the transcription factor *STAT3*, a pathway that synergizes with NF- $\kappa$ B in promoting cell survival<sup>9,10</sup>. The oncogenic aetiology of this JAK-*STAT3* signalling has not been elucidated.

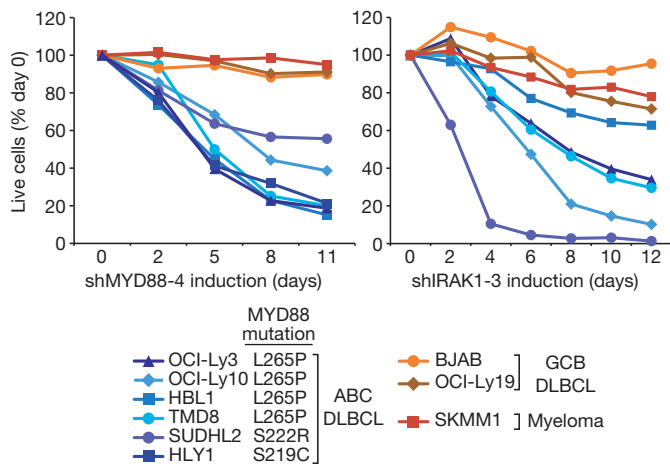
We conducted an RNA interference (RNAi) screen for genes that are required for proliferation and survival of lymphoma cell lines and identified three small hairpin RNAs (shRNAs) targeting *MYD88* that were toxic to two ABC DLBCL lines but not to two GCB DLBCL lines (Supplementary Fig. 1a). During normal immune responses, *MYD88* functions as a signalling adaptor protein that activates the NF- $\kappa$ B pathway after stimulation of toll-like receptors (TLRs) and receptors for IL-1 and IL-18 (refs 2, 3). *MYD88* coordinates the assembly of a multi-subunit signalling complex consisting of various members of the *IRAK* family of serine-threonine kinases<sup>11</sup>. The initial RNAi screen also identified two shRNAs targeting *IRAK1* as toxic for one or both of the ABC DLBCL lines, but not for GCB DLBCL lines. A subsequent screen identified additional *MYD88* and *IRAK1* shRNAs that were toxic to all five ABC DLBCL lines tested but had little effect on GCB DLBCL, Burkitt's lymphoma, mantle cell lymphoma and multiple myeloma lines (Supplementary Fig. 1a). Using shRNAs targeting the 3' untranslated regions of *MYD88* and *IRAK1*, which reduced expression of their respective proteins (Supplementary Fig. 1c), we showed that ABC DLBCL cells could be rescued from shRNA-mediated toxicity by coexpression of coding region cDNAs (*IRAK1*, Supplementary Fig. 1d; *MYD88*, see below). *MYD88* and *IRAK1* shRNAs displayed a time-dependent toxicity for ABC DLBCL lines and induced apoptosis, but had little effect on GCB DLBCL and myeloma lines (Fig. 1 and Supplementary Fig. 1b, e). Together these data establish that *MYD88* and *IRAK1* are required to maintain the viability of ABC DLBCL cells.

To comprehensively discover somatic mutations in ABC DLBCL, we used high-throughput resequencing of mRNA to search for sequence variants in four ABC DLBCL lines. In addition to known mutations in *CARD11* and *CD79B*, we identified a single nucleotide variant that changed a leucine residue at position 265 of the *MYD88* coding region to proline (L265P) in all four ABC DLBCL lines tested. This variant resides in the *MYD88* TIR domain, which interacts with TIR domains of various receptors during innate immune responses and also mediates homotypic interactions<sup>12,13</sup>.

To extend this finding, we resequenced the *MYD88* coding region in 382 lymphoma biopsy samples. The L265P mutation was by far the most

<sup>1</sup>Metabolism Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. <sup>2</sup>Bioinformatics and Molecular Analysis Section, Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>3</sup>Howard Hughes Medical Institute-National Institutes of Health Research Scholars Program, Bethesda, Maryland 20892, USA. <sup>4</sup>Biometric Research Branch, DCTD, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. <sup>5</sup>Department of Pathology, University of Würzburg, 97080 Würzburg, Germany. <sup>6</sup>Department of Clinical Pathology, Robert-Bosch-Krankenhaus, and Dr Margarete Fischer-Bosch Institute for Clinical Pharmacology, 70376 Stuttgart, Germany. <sup>7</sup>British Columbia Cancer Agency, Vancouver, British Columbia V5Z 4E6, Canada. <sup>8</sup>Department of Pathology, University of Arizona, Tucson, Arizona 85724, USA. <sup>9</sup>Southwest Oncology Group, 24 Frank Lloyd Wright Drive, Ann Arbor, Michigan 48106, USA. <sup>10</sup>Hospital Clinic, University of Barcelona, 08036 Barcelona, Spain. <sup>11</sup>Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA. <sup>12</sup>Pathology Clinic, Rikshospitalet University Hospital, N-0310 Oslo, Norway. <sup>13</sup>Institute for Cancer Research, Rikshospitalet University Hospital and Center for Cancer Biomedicine, Faculty Division of the Norwegian Radium Hospital, University of Oslo, N-0310 Oslo, Norway. <sup>14</sup>James P. Wilmot Cancer Center, University of Rochester School of Medicine, Rochester, New York 14642, USA. <sup>15</sup>Oregon Health and Science University, Portland, Oregon 97239, USA. <sup>16</sup>Cleveland Clinic Pathology and Laboratory Medicine Institute, Cleveland, Ohio 44195, USA. <sup>17</sup>Departments of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. <sup>†</sup>Present address: Division of Hematopoietic Stem Cell and Leukemia Research, City of Hope National Medical Center, Duarte, California 91010, USA.

\*These authors contributed equally to this work.



**Figure 1 | MYD88 is required for survival of ABC DLBCL cells.** *MYD88* and *IRAK1* shRNAs have selective toxicity for ABC DLBCL lines. Shown is the fraction of GFP<sup>+</sup>, shRNA-expressing cells relative to the GFP<sup>+</sup>, shRNA-negative fraction at the indicated times, normalized to the day 0 values.

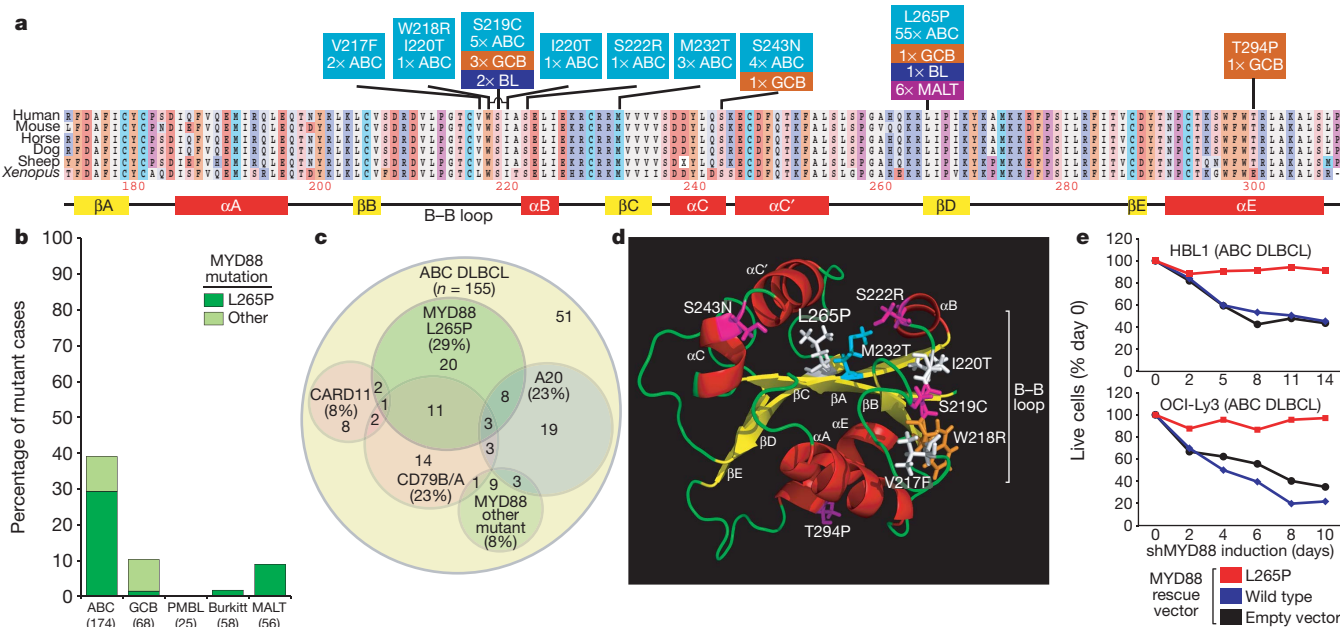
common variant observed, occurring in 29% of ABC DLBCL biopsies. By contrast, this mutation was rare or absent among DLBCLs of the GCB and PMBL subtypes and among Burkitt's lymphomas (Fig. 2b). Of note, *MYD88* L265P was also observed in 9% of gastric mucosa-associated lymphoid tissue (MALT) lymphomas. Most *MYD88* L265P mutations appeared heterozygous by sequencing, but six biopsy samples and one ABC DLBCL line (OCI-Ly3) were homozygous. By array-based comparative genomic hybridization<sup>14</sup>, 56% (15 of 27) of the ABC DLBCL cases with gain or amplification of the *MYD88* locus had the L265P mutation, compared to 29% (13 of 45) with wild-type *MYD88* copy number ( $P = 0.023$ ), indicating selection by the cancer cells for this mutant allele. A host of other, less common *MYD88* mutations were equally distributed among ABC and GCB DLBCL cases (Fig. 2a, b). Whereas most mutations were in the TIR domain, one mutation

(V52M) was in the death domain and two were between the death and TIR domains (S149G/I). Six ABC DLBCL lines had a *MYD88* mutation (Fig. 1), whereas all 14 GCB DLBCL lines tested were wild type. In 13 DLBCL cases for which matched germline DNA was available, the *MYD88* mutations (L265P, V217F, S219C, M232T, S243N, T294P) were confirmed to be somatically acquired. Overall, *MYD88* mutations were observed in 39% of ABC DLBCLs (Fig. 2b), establishing *MYD88* as among the most frequently altered genes in this malignancy.

The *MYD88* mutations partially overlapped with abnormalities in *CD79B/A*, *A20* and *CARD11* in ABC DLBCL tumours (Fig. 2c). Among cases with a *MYD88* L265P mutation, 34% had a coincident *CD79B/A* mutation whereas this overlap was significantly less common among ABC DLBCLs without a *CD79B/A* mutation (18%;  $P = 0.03$ ). These data raise the possibility of a functional interaction between the chronic active B-cell receptor signalling that is associated with *CD79B/A* mutations<sup>8</sup> and the signalling that is instigated by the *MYD88* L265P mutation. Some cases had *MYD88* L265P as well as a *CARD11* mutation, which strongly activates NF- $\kappa$ B, suggesting that the *MYD88* mutation confers additional biological attributes beyond NF- $\kappa$ B activation.

The location of the *MYD88* mutations within the three-dimensional structure of the *MYD88* TIR domain was both surprising and instructive (Fig. 2d). The L265P mutation occurs at a residue that is invariant in evolution and contributes to a  $\beta$ -sheet at the hydrophobic core of the domain. Another mutation, M232T, affects a methionine that is in an adjacent  $\beta$ -sheet and contacts the leucine affected by L265P. A cluster of mutations were in the 'B-B loop', an evolutionarily conserved region that mediates TIR domain interactions<sup>15</sup>. Two other mutations, S222R and S243N, alter an adjoining face of the TIR domain. Only one mutant affects the opposite side of the TIR domain (T294P), altering the conserved 'box 3' motif that is important in IL-1 signalling<sup>13</sup>.

To examine whether the *MYD88* mutants confer a gain or loss of function, we performed a complementation experiment in which we knocked down endogenous *MYD88* in ABC DLBCL lines and ectopically expressed wild-type or mutant *MYD88* coding regions. In ABC



**Figure 2 | MYD88 mutations in human lymphomas.** **a**, *MYD88* missense mutations in lymphoma biopsies and cell line models of ABC DLBCL (light blue), GCB DLBCL (orange), MALT lymphoma (purple) and Burkitt's lymphoma (BL; dark blue). Amino acid positions are shown according to protein accession NP\_002459. **b**, Frequencies of *MYD88* mutations in biopsy samples from different lymphoma subtypes. **c**, Overlap of *MYD88* mutations with other recurrent genetic alterations in ABC DLBCL tumour specimens. Genetic subsets were defined by somatic mutations and, in the case of the A20

subset, by homozygous deletion or epigenetic silencing. **d**, Location of *MYD88* mutations in the three-dimensional structure of the *MYD88* TIR domain.

**e**, Dependence of ABC DLBCLs on *MYD88* L265P. A 3'-UTR-directed *MYD88* shRNA was inducibly expressed in the indicated ABC DLBCL lines, which were stably transduced with rescue vectors expressing wild-type or L265P *MYD88* coding regions, or with an empty vector. Shown is the fraction of viable shRNA-expressing cells relative to the shRNA-negative fraction, normalized to day 0 values.

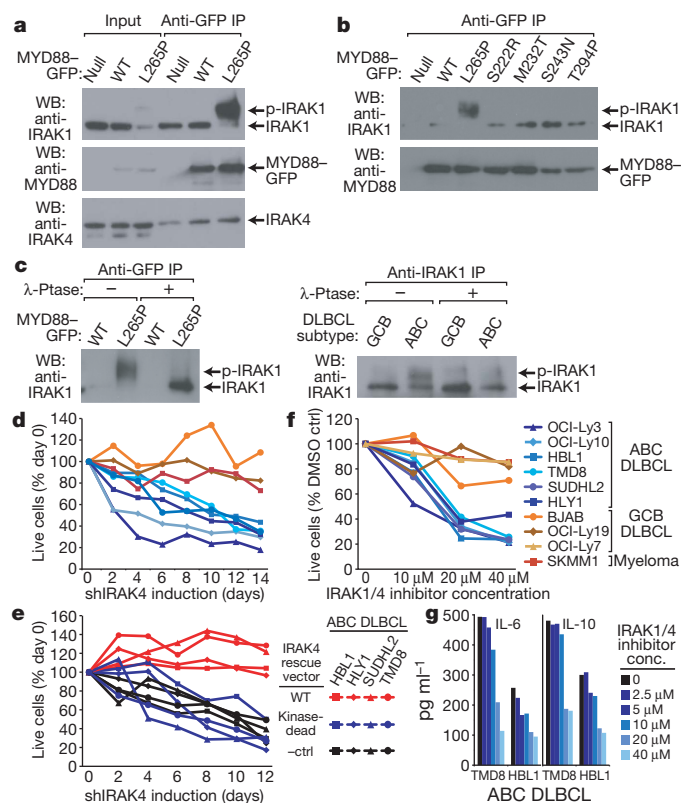
DLBCL lines harbouring an L265P mutation, *MYD88* L265P rescued the cells after *MYD88* knockdown, but wild-type *MYD88* was ineffective (Fig. 2e), although these *MYD88* isoforms were expressed equivalently (data not shown). Hence, these ABC DLBCLs are 'addicted' to the action of the L265P *MYD88* mutant, indicating that it is a gain-of-function driver mutation that confers a selective advantage during the evolution of ABC DLBCL tumours.

To assess the biochemical and functional consequences of the *MYD88* mutations, we fused green fluorescent protein (GFP) to *MYD88* and introduced the fusion proteins into DLBCL lines. Immunoprecipitation of *MYD88*-GFP with anti-GFP antibodies brought down IRAK1 and IRAK4, two kinases known to associate with *MYD88* upon TLR or IL-1 stimulation (Fig. 3a). During IL-1 signalling, IRAK1 becomes hyperphosphorylated by IRAK4, resulting in slowly migrating IRAK1 isoforms<sup>16</sup>. In cells bearing the *MYD88* L265P, a prominent, slow-migrating IRAK1 species co-immunoprecipitated with *MYD88* (Fig. 3a). By contrast, wild-type *MYD88* did not associate strongly with these IRAK1 isoforms nor did the other *MYD88* mutants tested (Fig. 3b). Treatment with  $\lambda$ -phosphatase collapsed the slow-migrating IRAK1 species into a single band, confirming that they are phosphorylated IRAK1 isoforms (Fig. 3c). Phosphorylation of endogenous IRAK1 was observed in an ABC DLBCL line with L265P but not in a GCB DLBCL line (Fig. 3c). Thus, the *MYD88* L265P mutant nucleates a signalling complex in ABC DLBCLs that includes phosphorylated IRAK1, consistent with a gain-of-function phenotype.

IRAK4 co-immunoprecipitated with *MYD88*, but it associated equivalently with wild-type and L265P *MYD88* (Fig. 3a). Knockdown of IRAK4 was toxic for ABC DLBCL lines but not for GCB DLBCL and myeloma lines (Fig. 3d and Supplementary Fig. 1c). Wild-type IRAK4 rescued ABC DLBCL lines after *IRAK4* shRNA induction, but a kinase-dead IRAK4 isoform could not (Fig. 3e), despite equivalent expression (data not shown). By contrast, IRAK1 kinase activity was not required for the survival of ABC DLBCL cells (Supplementary Fig. 1d). A selective small-molecule inhibitor of IRAK1 and IRAK4 kinase activity<sup>17</sup> killed ABC DLBCL lines but not GCB DLBCL and myeloma lines (Fig. 3f). Together, these findings demonstrate that ABC DLBCLs rely upon IRAK4 kinase activity to transduce signals from *MYD88* L265P that promote survival.

To investigate signalling pathways that are engaged by *MYD88* L265P, we knocked it down in an ABC DLBCL line and profiled the ensuing gene expression changes (Supplementary Table 1 and Supplementary Fig. 2). We identified 285 genes that were down-modulated after *MYD88* knockdown, and searched for overlap between this *MYD88* signature and previously defined gene expression signatures<sup>18</sup> (Supplementary Table 2). The most significantly enriched signature reflects NF- $\kappa$ B signalling in ABC DLBCL (44 $\times$  enrichment,  $P = 2.4 \times 10^{-130}$ ). This signature was also inhibited after *IRAK1* knockdown (Supplementary Fig. 3), indicating that IRAK1 mediates NF- $\kappa$ B activation by *MYD88* L265P. To compare the ability of wild-type and mutant *MYD88* isoforms to activate NF- $\kappa$ B, we expressed them as GFP fusion proteins in a GCB DLBCL line with little endogenous NF- $\kappa$ B activity. Whereas wild-type *MYD88* activated an NF- $\kappa$ B-dependent reporter modestly, L265P had strong activity, as did M232T and S243N, whereas S222R and T294P had an intermediate effect (Fig. 4b). At all *MYD88* expression levels, L265P was superior to wild-type *MYD88* in upregulating CD83, a previously established NF- $\kappa$ B target in this system<sup>5</sup> (Fig. 4c). Other *MYD88* mutants induced CD83 to varying degrees but all were more active than wild-type *MYD88*. Thus, mutant *MYD88* isoforms can contribute to the constitutive NF- $\kappa$ B activation that typifies ABC DLBCL<sup>19</sup>.

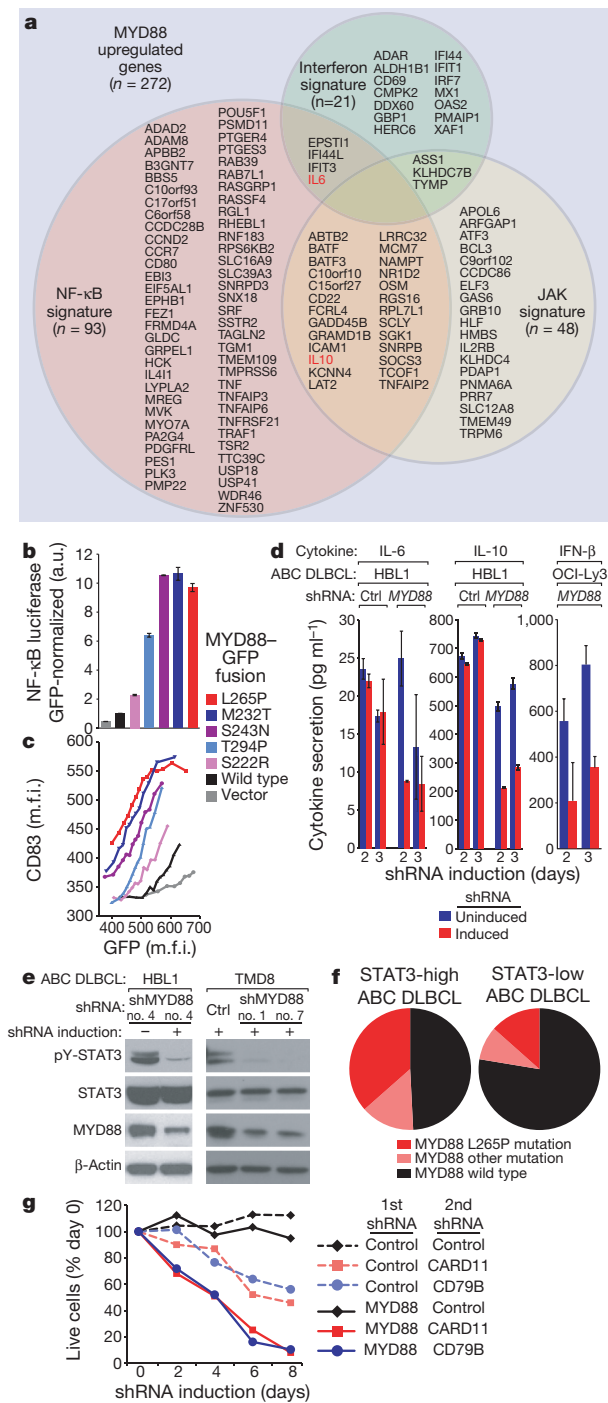
A signature of JAK kinase signalling in ABC DLBCL overlapped significantly with the *MYD88* signature (Fig. 4a; 14 $\times$  enrichment,  $P = 9.6 \times 10^{-39}$ ) and with IRAK1-regulated genes (Supplementary Fig. 3b). This was notable because autocrine secretion of IL-6 and IL-10 drives JAK-STAT3 signalling in a subset of ABC DLBCLs<sup>9</sup>. *MYD88* knockdown significantly diminished the secretion of IL-6 and



**Figure 3** | *MYD88* mutations are gain-of-function. **a**, An altered IRAK1 isoform associated with *MYD88* L265P. The GCB DLBCL line BJAB was transduced with GFP-tagged wild-type (WT) or L265P *MYD88*, or with an empty vector (null). Anti-GFP immunoprecipitates (IP) and input lysates were examined by immunoblotting for IRAK1, IRAK4 and *MYD88*. **b**, Preferential association of an altered IRAK1 isoform with *MYD88* L265P. BJAB cells were transduced with the indicated GFP-tagged *MYD88* isoforms and examined as in **a**. **c**, *MYD88* L265P associates with phosphorylated IRAK1. Top panel: BJAB cells were transduced with the indicated GFP-tagged *MYD88* isoforms. Anti-GFP immunoprecipitates were treated with  $\lambda$ -phosphatase as indicated and examined by immunoblotting for IRAK1 or *MYD88*. Bottom panel: anti-IRAK1 immunoprecipitates from HBL1 (ABC) or BJAB (GCB) cells were treated with  $\lambda$ -phosphatase as indicated and examined by immunoblotting for IRAK1. **d**, Toxicity of *IRAK4* shRNAs for ABC DLBCLs. The indicated lines were transduced with retroviruses expressing *IRAK4* shRNA and the relative number of shRNA<sup>+</sup> cells is plotted versus time after shRNA induction, normalized to day 0. Data are representative of experiments with three different *IRAK4* shRNAs. **e**, IRAK4 kinase activity is required for ABC DLBCL survival. The indicated ABC DLBCL lines were transduced with retroviruses expressing wild-type or kinase-dead IRAK4 isoforms, or with an empty vector (ctrl). The survival of cells after induction of an *IRAK4* shRNA is shown. **f**, A small-molecule IRAK1/4 kinase inhibitor is selectively lethal for ABC DLBCLs. Viability of the indicated lines was measured after treatment for 3 days with various inhibitor concentrations and normalized to DMSO-treated cells. **g**, IRAK4 kinase activity regulates IL-6 and IL-10 secretion. The indicated cytokines were measured in the supernatant of ABC DLBCL lines after treatment for 24 h with various concentrations of the IRAK1/4 inhibitor.

IL-10 as well as the phosphorylation of STAT3 in several ABC DLBCL lines (Fig. 4d, e and Supplementary Fig. 1f). IL-6 and IL-10 secretion was also blocked by the IRAK1/4 kinase inhibitor, indicating that IRAK4 links *MYD88* L265 signalling to the expression of these cytokines (Fig. 3g). Previous work identified a 'STAT3-high' subgroup of ABC DLBCL tumours with autocrine IL-6/IL-10 signalling and STAT3 phosphorylation, which was missing in a 'STAT3-low' subgroup<sup>9</sup>. The *MYD88* L265P mutation was significantly more common in the STAT3-high subgroup (37%) than in the STAT3-low subgroup (13%) ( $P = 0.0036$ ), and other *MYD88* mutations were modestly enriched among STAT3-high cases as well (Fig. 4f). These data indicate that





MYD88 mutations contribute to JAK–STAT3 signalling in ABC DLBCL tumours.

The MYD88 signature included a number of genes that are induced by type I interferon (Fig. 4a; 7× enrichment,  $P = 4.3 \times 10^{-10}$ ), which is intriguing given that MYD88 signalling can induce interferon (IFN)-β production by innate immune cells. IFN-β was measurable in the supernatant of the OCI-Ly3 ABC DLBCL line, and MYD88 knockdown diminished its secretion (Fig. 4d). MYD88 knockdown decreased IFN-β mRNA levels in the HBL1 line (Supplementary Fig. 2), although IFN-β secretion was below the detection limit. Future work should address whether the secretion of immunomodulatory cytokines such as IFN-β, IL-6 and IL-10 influences immune cells in the microenvironment of ABC DLBCL tumours.

Given the pleiotropic action of MYD88 L265P, we investigated whether MYD88 signalling cooperates with B-cell receptor signalling

**Figure 4 | MYD88 mutants activate NF-κB and cytokine signalling.** **a**, Venn diagram of genes affected by MYD88 knockdown in the HBL1 ABC DLBCL line, grouped according to membership in gene expression signatures. **b**, MYD88 mutants constitutively activate NF-κB. BJAB cells co-expressing the indicated MYD88–GFP mutants and a NF-κB-driven luciferase reporter construct were assayed for luciferase activity, which was normalized to the expression levels of each MYD88–GFP isoform. a.u., arbitrary units. **c**, Correlation of MYD88 protein levels with CD83 expression. BJAB cells bearing GFP-tagged MYD88 isoforms were assessed for CD83 and GFP expression. Cells were assigned to equally sized bins based on their GFP levels, and the mean fluorescence intensity (m.f.i.) of CD83 in each bin was plotted. **d**, MYD88 knockdown decreases cytokine secretion in ABC DLBCL. MYD88 or control (ctrl) shRNAs were induced in ABC DLBCL lines, and the indicated cytokines were measured in the supernatant over time. **e**, STAT3 phosphorylation in ABC DLBCL depends on MYD88. MYD88 or control (ctrl) shRNAs were induced in ABC DLBCL lines, and cells were assessed by immunoblotting for phosphorylated STAT3 (pY-STAT3), total STAT3, MYD88 and β-actin. **f**, Preferential association of MYD88 mutant isoforms with the STAT3-high subgroup of ABC DLBCL tumours. See text for details. **g**, MYD88 and B-cell-receptor signalling pathways cooperate to maintain ABC DLBCL survival. OCI-Ly10 ABC DLBCL cells were first transduced with retroviruses expressing MYD88 or control shRNAs and then infected with retroviruses expressing CD79A, CARD11, or control shRNAs along with GFP. The relative viability of GFP<sup>+</sup> cells is plotted, normalized to day 0 values. All error bars are s.e.m. ( $n = 3$ ).

to maintain ABC DLBCL survival. Knockdown of MYD88 enhanced the killing of an ABC DLBCL line with chronic active BCR signalling by CD79B or CARD11 shRNAs (Fig. 4g). This finding indicates that MYD88 and B-cell receptor signalling provide non-redundant survival signals to ABC DLBCL cells, in keeping with the fact that some ABC DLBCL tumours harbour MYD88 L265P as well as CD79B or CARD11 mutations (Fig. 2c).

Our genetic and functional data establish a new oncogenic pathway in lymphomagenesis. Somatically acquired MYD88 mutations in ABC DLBCL promote NF-κB and JAK–STAT3 signalling, which mediate cell survival in this lymphoma type<sup>9,19</sup>. MYD88 L265P was the most biologically potent mutant and was unique in its ability to coordinate a stable signalling complex containing phosphorylated IRAK1, which probably accounts for its high recurrence among lymphomas. MYD88 L265P also genetically links MALT lymphoma with ABC DLBCL, two lymphoma subtypes that share other oncogenic features<sup>7,14,20–22</sup>. Other MYD88 mutations may also be drivers of lymphomagenesis given their recurrent nature and ability to activate NF-κB. From a therapeutic standpoint, the signalling complex coordinated by MYD88 L265P represents an enticing target. Our study also provides a genetic method to identify patients whose tumours may depend upon MYD88 signalling and who may therefore benefit from therapies targeting IRAK4 alone or in combination with agents targeting the B-cell receptor<sup>8</sup>, NF-κB<sup>23,24</sup> or JAK–STAT3 pathways<sup>9</sup>.

## METHODS SUMMARY

RNAi screens, doxycycline-inducible shRNA expression and shRNA toxicity assays were described previously<sup>25</sup>. RNA interference screening results are listed in Supplementary Table 3. The sequences of individual shRNAs described in the figures are given in Supplementary Table 4. Gene expression profiling data have been submitted to GEO under accession number GSE22900.

Pre-treatment tumour biopsies were obtained from patients with *de novo* DLBCL, Burkitt's lymphoma and gastric MALT lymphoma. Samples were analysed as per a protocol approved by the National Cancer Institute Institutional Review Board. Assignment of DLBCL specimens to subtypes was performed as described<sup>1</sup>. High-throughput RNA sequencing was accomplished using an Illumina GAIIx instrument.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 21 July; accepted 17 November 2010.

Published online 22 December 2010.

1. Lenz, G. *et al.* Stromal gene signatures in large-B-cell lymphomas. *N. Engl. J. Med.* **359**, 2313–2323 (2008).

2. Iwasaki, A. & Medzhitov, R. Regulation of adaptive immunity by the innate immune system. *Science* **327**, 291–295 (2010).
3. Ishii, K. J. & Akira, S. Innate immune recognition of, and regulation by, DNA. *Trends Immunol.* **27**, 525–532 (2006).
4. Lenz, G. & Staudt, L. M. Aggressive lymphomas. *N. Engl. J. Med.* **362**, 1417–1429 (2010).
5. Lenz, G. *et al.* Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science* **319**, 1676–1679 (2008).
6. Compagno, M. *et al.* Mutations of multiple genes cause deregulation of NF- $\kappa$ B in diffuse large B-cell lymphoma. *Nature* **459**, 717–721 (2009).
7. Kato, M. *et al.* Frequent inactivation of A20 in B-cell lymphomas. *Nature* **459**, 712–716 (2009).
8. Davis, R. E. *et al.* Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88–92 (2010).
9. Lam, L. T. *et al.* Cooperative signaling through the signal transducer and activator of transcription 3 and nuclear factor- $\kappa$ B pathways in subtypes of diffuse large B-cell lymphoma. *Blood* **111**, 3701–3713 (2008).
10. Ding, B. B. *et al.* Constitutively activated STAT3 promotes cell proliferation and survival in the activated B-cell subtype of diffuse large B-cell lymphomas. *Blood* **111**, 1515–1523 (2008).
11. Lin, S. C., Lo, Y. C. & Wu, H. Helical assembly in the MyD88–IRAK4–IRAK2 complex in TLR/IL-1R signalling. *Nature* **465**, 885–890 (2010).
12. Xu, Y. *et al.* Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* **408**, 111–115 (2000).
13. Li, C., Zienkiewicz, J. & Hawiger, J. Interactive sites in the MyD88 Toll/interleukin (IL) 1 receptor domain responsible for coupling to the IL1 $\beta$  signaling pathway. *J. Biol. Chem.* **280**, 26152–26159 (2005).
14. Lenz, G. *et al.* Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc. Natl Acad. Sci. USA* **105**, 13520–13525 (2008).
15. Jiang, Z. *et al.* Details of Toll-like receptor:adapter interaction revealed by germ-line mutagenesis. *Proc. Natl Acad. Sci. USA* **103**, 10961–10966 (2006).
16. Yamin, T. T. & Miller, D. K. The interleukin-1 receptor-associated kinase is degraded by proteasomes following its phosphorylation. *J. Biol. Chem.* **272**, 21540–21547 (1997).
17. Powers, J. P. *et al.* Discovery and initial SAR of inhibitors of interleukin-1 receptor-associated kinase-4. *Bioorg. Med. Chem. Lett.* **16**, 2842–2845 (2006).
18. Shaffer, A. L. *et al.* A library of gene expression signatures to illuminate normal and pathological lymphoid biology. *Immunol. Rev.* **210**, 67–85 (2006).
19. Davis, R. E., Brown, K. D., Siebenlist, U. & Staudt, L. M. Constitutive nuclear factor  $\kappa$ B activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J. Exp. Med.* **194**, 1861–1874 (2001).
20. Dierlamm, J. *et al.* Characteristic pattern of chromosomal gains and losses in marginal zone B cell lymphoma detected by comparative genomic hybridization. *Leukemia* **11**, 747–758 (1997).
21. Novak, U. *et al.* The NF- $\kappa$ B negative regulator TNFAIP3 (A20) is inactivated by somatic mutations and genomic deletions in marginal zone B-cell lymphomas. *Blood* **113**, 4918–4921 (2009).
22. Staudt, L. M. Oncogenic activation of NF- $\kappa$ B. *Cold Spring Harb. Perspect. Biol.* **2**, a000109 (2010).
23. Milhollen, M. A. *et al.* MLN4924, a NEDD8-activating enzyme inhibitor, is active in diffuse large B-cell lymphoma models: rationale for treatment of NF- $\kappa$ B-dependent lymphoma. *Blood* **116**, 1515–1523 (2010).
24. Lam, L. T. *et al.* Small molecule inhibitors of I $\kappa$ B kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clin. Cancer Res.* **11**, 28–40 (2005).
25. Ngo, V. N. *et al.* A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, an NCI SPECS grant (U01-CA 114778), and by the Foundation for NIH, through a gift from the Richard A. Lauderbaugh Memorial Fund. This study was conducted under the auspices of the Lymphoma/Leukemia Molecular Profiling Project (LLMPP). R.S. is supported by the Dr Mildred Scheel Stiftung für Krebsforschung (Deutsche Krebshilfe). P.R. was an HHMI-NIH Research Scholar. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Maryland (<http://biowulf.nih.gov>). We thank D. Staudt for discussions, K. Meyer for help with the GEO submission, and X. Li for IRAK1 plasmids. We are grateful to B. Tran and the Center for Cancer Research Sequencing Facility for implementation of next generation RNA sequencing.

**Author Contributions** V.N.N., R.M.Y., R.S., S.J., K.-H.L., H.K. and A.L.S. designed and performed experiments. W.X., Y.Y. and H.Z. performed experiments. W.X., G.W. and J.P. analysed data. A.R., H.K.M.-H., G.O., R.D.G., J.M.C., L.M.R., E.C., E.S.J., J.D., E.B.S., R.I.F., R.M.B., R.R.T., J.R.C., D.D.W. and W.C.C. supplied patient samples and reviewed pathological and clinical data. L.M.S. designed and supervised research and wrote the manuscript.

**Author Information** Gene expression profiling data have been submitted to GEO under accession number GSE22900. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.M.S. ([lstaudt@mail.nih.gov](mailto:lstaudt@mail.nih.gov)).

## METHODS

**Cell lines.** Cell lines were cultured in RPMI 1640 medium supplemented with penicillin/streptomycin and 10% fetal bovine serum or, for the OCI series of cell lines, Iscove's medium with 20% fresh human plasma. Cells were maintained in a humidified, 5% CO<sub>2</sub> incubator at 37 °C. All cell lines were engineered to express an ecotropic retroviral receptor and the bacterial tetracycline repressor as previously described<sup>25</sup>.

**Retroviral vectors and retroviral transduction.** A previously described retroviral vector, pRSMX<sup>25</sup>, was used to express shRNA for the library screen. A modified version of this vector, pRSMX-PG, in which the puromycin selectable marker was fused with the green fluorescence protein (GFP), was used to co-express shRNA and GFP for shRNA toxicity assay. Retroviral transduction was performed by transfecting the retroviral vector and a mixture of helper plasmids for a mutant ecotropic envelope and *gag* and *pol* into 293T cells using Lipofectamine 2000 (Invitrogen). Retroviral supernatants were harvested 48 h after transfection and were used to transduce ecotropic receptor-expressing target cells by centrifugation at 2,500 r.p.m. for 1.5 h in the presence of 8 µg ml<sup>-1</sup> polybrene.

**shRNA library screening.** Pools of shRNA library were screened as previously described<sup>25</sup>. Briefly, pools of roughly 1,000 shRNA expressing retroviral vectors were used to transduce target cell lines. After puromycin selection, stable integrants were induced to express shRNA by doxycycline (20 ng ml<sup>-1</sup>). Parallel uninduced cultures were kept as control. After 3 weeks of shRNA induction, genomic DNA from both uninduced and induced cultures were harvested. shRNA-associated bar code sequences from the genomic DNA were PCR amplified and *in vitro* transcribed, as described<sup>25</sup>. The transcribed RNA products were labelled fluorescently with either Cy5 (induced) or Cy3 (uninduced) using the Universal Linkage System (Amersham Biosciences) and hybridized onto microarrays containing DNA oligonucleotides complementary to the bar code sequences, as described<sup>25</sup>. Each bar code experiment was performed in quadruplicate, and the microarray results for each bar code were averaged. The complete screening results are presented in Supplementary Table 3, which includes some data that have been previously published<sup>8,25</sup>.

**shRNA sequences.** The shRNA sequences used in individual experiments are listed in Supplementary Table 4.

**shRNA toxicity and complementation assays.** shRNA toxicity was assayed as described<sup>25</sup>. Briefly, the pRSMX-PG vector that co-expresses shRNA and GFP was transduced into lymphoma and multiple myeloma cell lines. Two days after retroviral transduction, doxycycline was added to induce shRNA expression. The fraction of GFP<sup>+</sup>, shRNA-expressing cells relative to the GFP<sup>+</sup>, shRNA<sup>-</sup> fraction was monitored over various time points by flow cytometry and plotted against the same GFP<sup>+</sup>, shRNA-expressing fraction on day 0 of doxycycline induction. The reduction of the GFP<sup>+</sup>, shRNA-expressing fraction at later time points indicates shRNA toxicity. Complementation studies were performed in the DLBCL cell line that harbours the MYD88 L265P mutation. HBL1 cells were transduced with retroviral vectors that co-express GFP and shRNA targeting the 3' UTR of *MYD88* (or *IRAK1*). shRNA-transduced cells were subsequently infected with retroviruses co-expressing wild-type or mutant *MYD88* (or *IRAK1*) coding regions and mouse CD8 (Lyt2). The cell fraction positive for GFP and CD8 (using anti-mouse CD8a, BD Pharmingen) was monitored over time by flow cytometry as above. TMD8 and OCI-Ly3 cells were first transduced with retroviruses co-expressing wild-type or mutant *MYD88* (or *IRAK1*) coding regions and mouse CD8 (Lyt2) and enriched for Lyt2 expression with magnetic beads. Enriched cells were subsequently infected with retroviral vectors that co-express GFP and an shRNA targeting the 3' UTR of *MYD88*. The cell fraction positive for GFP and CD8 was monitored over time by flow cytometry.

For the *IRAK4* complementation assay, HBL1, HLY1, TMD8 and SUDHL2 lymphoma cell lines were first retrovirally transduced with either vLyt2 empty vector, or vLyt2 vector expressing wild-type or kinase-dead *IRAK4* (K213A/K214A). The infected cells were later enriched using the CD8 microbeads method (Miltenyi Biotec) according to manufacturer's protocol. The enriched cells were then infected with retroviral vector pRSMX-PG expressing either a control shRNA or an *IRAK4* shRNA. Doxycycline was added 2 days after infection to induce shRNA expression. The fraction of GFP-positive cells was monitored over time by FACS analysis and the decline of GFP hence shRNA expressing cells indicates toxicity.

**Synergistic toxicity of MYD88 and either CARD11 or CD79A knockdown.** OCI-Ly10 cells were first infected with either pRSMX-puro empty vector or pRSMX-puro vector encoding *MYD88* shRNA sequence (5'-GTACCAGTATTTATCTACTTA-3'). Two days after infection the two cell lines were selected using 1 µg ml<sup>-1</sup> puromycin. The selected cells were then retrovirally infected with pRSMX-PG encoding either a scramble control, shRNA sequence against *CARD11* (target sequence 5'-GGGGTGTGTACCAGGCTATGA-3') or *CD79A* (target sequence 5'-GGGGCTTCCTTAGTCATATTC-3'). Doxycycline was

added 2 days after infection to induce shRNA expression. The fraction of GFP-positive cells was monitored over time by FACS analysis and the decline of GFP hence shRNA expressing cells indicates toxicity.

**High-throughput RNA sequencing/PCR amplification/Sanger sequencing.** The standard Illumina pipeline for RNA-seq was used, using paired-end 75-base-pair runs with each sample run in one sequencing lane, yielding ~20 million reads per sample. Sequences were mapped back to both RefSeq and Ensemble transcript models using the BWA algorithm<sup>26</sup>, yielding a median resequencing coverage of 10×. Single nucleotide variants (SNVs) were reported that deviated from the human reference genome sequence, were observed in both sequencing directions, represented >20% of the resequencing coverage at a particular base pair position, and were not known SNPs in the dbSNP database of NCBI. A total of 52,160 putative SNVs was detected in the four ABC DLBCL cell lines studied. Sequences have been submitted to the NCBI short sequence archive under accession SRP003192. On the basis of the criteria above, all SNVs that are not represented in publically available databases of single nucleotide polymorphisms (SNPs) are listed in Supplementary Table 5. Except for the MYD88 mutations, other SNVs in this table have not been validated by independent means.

Sanger sequencing of MYD88 was accomplished with the following primers: MYD88-Full-F, 5'-GACCTCTCCAGATCTCAAAAGGCAGATTCC-3' (PCR amplification and sequencing, exon 1); MYD88-Full-R, 5'-GCAGAAGTACATGGACAGGCAGACAGATAC-3' (PCR amplification and sequencing, exon 5); MYD88-Seq-E1R, 5'-TCTCTCCATGGGAGACAGGATGCTG-3' (sequencing exon 1); MYD88-Seq-E2F, 5'-TGGGTAAGAGGTAGGCACTCCAG-3' (sequencing exon 2); MYD88-Seq-E2R, 5'-GCCCATCTGCTTCAACACCCATGC-3' (sequencing exon 2); MYD88-Seq-E3F, 5'-AAGCCTTCCCATGGAGCTCTGACAC-3' (sequencing exon 3); MYD88-Seq-E3R, 5'-GCTAGGAGGAGATGCCAGTATCTG-3' (sequencing exon 3); MYD88-Seq-E4F, 5'-ACTAAGTTGCACAGGACCTGCAGC-3' (sequencing exon 4); MYD88-Seq-E4R, 5'-ATCCA GAGGCCACCTACACATTC-3' (sequencing exon 4); MYD88-Seq-E5F, 5'-GTTGTTAACCTGGGGTTGAAG-3' (sequencing exon 5).

For 155 cases of ABC DLBCL, analysis of CARD11, CD79B, and A20 by Sanger sequencing was performed as described<sup>5,8,27</sup>.

A20 was declared epigenetically silenced if the expression level in a case was more than 2 standard deviations below the mean of other ABC DLBCL cases, based on previous gene expression profiling data<sup>1</sup>. Deletion of the A20 locus (*TNFAIP3*) was analysed by quantitative PCR using primers to amplify exon 3 and exon 6, as described<sup>27</sup>, and compared to a reference gene, *CHMP4A*, that is not subject to copy number alterations in ABC DLBCL. A20 was declared deleted if one or both of the A20 PCR products had an estimated copy number that was more than 3 standard deviations below the average of 9 normal control DNA samples. The following quantitative PCR primers were used: CHMP4A-F, 5'-CTGAAGGGAGGAGGGGTTTCATTC-3' (qPCR control gene); CHMP4A-R, 5'-CTTGGGTGTTCTTCTGGCCAGTC-3' (qPCR control gene); A20-E3F, 5'-ACCTTTGCTGGGTCTTACATGCAG-3' (qPCR A20); A20-E3R, 5'-TATGCCACCATGGAGCTCTGTTAG-3' (qPCR A20); A20-E6F, 5'-TGAGATCTACTTACCTATGGCCTTG-3' (qPCR A20); A20-E6R, 5'-TCAGGTGGCTGAGTTAAAGACAG-3' (qPCR A20).

**Expression vectors and cDNA mutagenesis.** The expression vector, vLyt2-MYD88-EGFP, encoding carboxy terminus EGFP-tagged MYD88 was constructed by three-way ligation of PCR-generated MYD88 and EGFP products into the pBMN-IRES-Lyt2 vector (provided by G. Nolan). The restriction site SalI was included in the MYD88 PCR reverse primer and the EGFP PCR forward primer to facilitate the ligation between MYD88 and EGFP. MYD88 or EGFP PCR products were generated using the following primer pairs: 5'-TAGTAGGGATCCG CCGCCACCATGCGACCGACCGCGCTGA-3' (MYD88), 5'-TAGTAGGTC GACGGGAGGGACAAGGCCTTGGC-3' (MYD88), 5'-TAGTAGGTCGACATGGTGAGCAAGGGCGGAGGAG-3' (EGFP), 5'-TAGTAGGCGGCCGCTTACTGTACAGCTCGTCCAT-3' (EGFP).

The expression vector vLyt2-AU1-MYD88 encoding amino terminus AU1-tagged MYD88 was constructed by inserting PCR-generated MYD88 into the pBMN-IRES-Lyt2 vector. MYD88 PCR product was generated using the following primers: 5'-TAGTAGGGATCCGCGCCGACCATGGCCGCGGCCGGGGCCGGGC-3' and 5'-TAGTAGGCGGCCGCTCACTTGTATCGTCGCTCTTGTAGTCGC TCTGAAATTCATCACTTTC-3'.

*IRAK1* expression vectors were similarly created by inserting PCR-generated *IRAK1* cDNAs (from pIND-*IRAK1* wild type and K239A kinase-dead templates, a gift from X. Li) into the pBMN-IRES-Lyt2 vector, using the following PCR primers: 5'-TAGTAGCTCGAGGCGGCCACCATGGCCGCGGCCGGGGCCGGGC-3' and 5'-TAGTAGGCGGCCGCTCACTTGTATCGTCGCTCTTGTAGTCGC TCTGAAATTCATCACTTTC-3'.

*IRAK4* expression vectors were generated by inserting PCR-generated *IRAK4* cDNA (from a template obtained from the Dana-Farber/Harvard Cancer Center



DNA Resource Core) into the pBMN-IRES-Lyt2 vector, using the following primers: 5'-TAGTAGGGATGGGCGCCACCATGGACACATACCGCTACATCAACAAACCCATAACACCATCA-3' and 5'-TAGTAGGCGGCGGCTCAAGAAGCTGTCATCTCTTGCAG-3'.

MYD88 mutants were created with the Phusion site-directed mutagenesis kit (New England BioLabs), using either vLyt2-MYD88-EGFP or vLyt2-AU1-MYD88 vector as templates. All cDNA inserts from PCR cloning and site-directed mutagenesis were verified by sequencing. The MYD88 mutagenesis primers used were the following: L265P forward P-CATCAGAAGCGACCGATCCCCATCAAG and L265P reverse P-GGCACCTGGAGAGAGGCTGAGTGCAGAA; M232T forward P-AGGTGCCGCCGACGGTGGTGGTTGTC and M232T reverse P-CTTTTCGATGAGCTCACTAGCAATAGA; S243N forward P-GAT TACCTGCAGAACAAAGGAATGTGAC and S243N reverse P-ATCAGAGACA ACCACCACCATCCGG; T294P forward P-ACCAACCCCTGCCCCAATCT TGGTTC and T294P reverse P-GTAGTCGACAGCAGTGATGAACCTCAG; S222R forward P-GGTCTATTGCTAGGGAGCTCATCGAAA and S222R reverse P-AGACACAGGTGCCAGGCAGGACATCGC.

The IRAK4 kinase-dead mutant was generated similarly using the following mutagenesis primers: K213A/K214A forward P-ACTGTGGCAGTGGCGGCG CTTGACGCAATG and K213A/K214A reverse P-TGTGTTATTACGTAGC CTTTATATACA.

**MYD88 co-immunoprecipitation.** BJAB cells were retrovirally transduced with various MYD88-GFP fusion constructs co-expressing a Lyt2 surface marker. Cells were enriched for Lyt2 expression using anti-Lyt2 magnetic beads (Invitrogen, 114.47D) following the manufacturer's instructions. Enriched cells were lysed at  $10^7$  cells per ml in RIPA buffer (0.5% Triton X-100, 0.5% deoxycholate, 0.05% SDS, 10 mM Tris, pH 8.0, 50 mM NaCl, 10 mM EDTA, 1 mM  $\text{Na}_3\text{VO}_4$ , 30 mM pyrophosphate, 10 mM glycerophosphate, 1 mM AEBSF, 0.02 U  $\text{ml}^{-1}$  aprotinin and 0.01%  $\text{Na}_2\text{S}_2\text{O}_8$ ) for 10 min on ice. Lysates were cleared by centrifuging for 20 min at 14,000g at 4 °C. MYD88-GFP constructs were immunoprecipitated with washed anti-GFP agarose beads (Chromotek) for 30 min at 4 °C, then washed 3–4 times in 1× RIPA buffer. For  $\lambda$ -phosphatase treatment, the agarose beads were washed two additional times in 10 mM Tris, pH 8.0 with 50 mM NaCl to remove EDTA and phosphates inhibitors.  $\lambda$ -phosphatase (New England Biolabs) treatment was done according to the manufacturer's instructions. Reactions were quenched by the addition of 2× lamellae sample buffer followed by boiling. Samples were separated on 10% polyacrylamide gels and transferred to Immobilon-P PVDF membranes (Millipore) for western blot analysis. Antibodies used for immunoblotting were anti-IRAK1 rabbit polyclonal (Santa Cruz Biotech), anti-IRAK4 rabbit polyclonal (Cell Signaling Technologies) and anti-MYD88 rabbit monoclonal (Cell Signaling Technologies).

**NF- $\kappa$ B reporter assay.** BJAB cells retrovirally expressing MYD88-GFP constructs (see above) were transduced with lentiviral particles containing a NF- $\kappa$ B firefly luciferase reporter construct by following the manufacturer's instructions (SA Biosciences). Firefly luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's instructions. Luminescence from equivalent amounts of lysate was read in triplicate on a Microtiter Plate Luminometer (Dyn-Ex Technologies). All readings were normalized to the mean fluorescence intensity of MYD88-GFP expression for each MYD88 mutant as determined by FACS analysis on a FACScalibur flow cytometer (Becton Dickinson).

**Western blotting.** Cells were lysed in lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% Triton X-100, 1% NP-40, 2 mM EDTA) supplemented with Complete Protease Inhibitor Cocktail Tablets (Roche) and phosphatase inhibitors (Sigma) for 30 min. Lysates were cleared by centrifugation at 15,000g at 4 °C for 10 min and protein concentrations were determined by BCA protein assay (Pierce). 80–100  $\mu$ g of lysates were subjected to electrophoresis through a 4–12% Bis-Tris gel (Invitrogen) and immobilized on the nitrocellulose membranes. Proteins were detected using the following antibodies: MYD88, IRAK4,  $\beta$ -actin, STAT3 and p-STAT3 (Y705) (Cell Signaling Technology).

**IRAK1 immunoprecipitation.** Cells were lysed at  $10^7$  cells per ml in RIPA buffer as described above. Lysates were pre-cleared with protein-A agarose beads (Pierce) before incubation with 1  $\mu$ g  $\text{ml}^{-1}$  of anti-IRAK1 polyclonal antibody (Santa Cruz Biotech, H-273) for 2 h on ice. Protein-A agarose beads were added to lysates and

rotated for 1 h at 4 °C, then washed three times with 1× RIPA buffer.  $\lambda$ -phosphatase treatment was performed as described above. Samples were separated on 10% polyacrylamide gels and transferred to Immobilon-P PVDF membranes (Millipore) for western blot analysis.

**Cytokine measurement.** The culture medium of cells transduced with inducible shRNAs was replaced with fresh medium plus doxycycline, and the concentrations of IL-6, IL-10 or IFN- $\beta$  in culture supernatants at the indicated times were measured by ELISA (R&D Systems). Alternatively, unmanipulated lymphoma cells lines were placed into fresh media with the addition of the IRAK1/4 inhibitor (EMD chemicals) and assessed for cytokines as above. The results were normalized to live cell numbers, and are representative of at least two independent experiments.

**Apoptosis measurements.** HBL1 cells were retrovirally transduced with either control or MYD88-specific shRNAs, as described above. shRNA expression was induced with doxycycline and cells were evaluated for apoptosis on 2, 3 and 4 days after induction. To measure apoptosis, cells were first fixed for 10 min with 1.5% paraformaldehyde, centrifuged and then fixed and permeabilized in cold methanol overnight. Methanol-fixed cells were washed three times with FACS buffer (PBS with 1% FBS) and stained with PE rabbit anti-active caspase 3 (BD Pharmingen) and Alexa Fluor 647 mouse anti-cleaved PARP (Asp 214) (BD Pharmingen) for 20 min at room temperature in the dark, followed by an additional wash with FACS buffer. Stained cells were subjected to FACS analysis (FACScalibur, BD) and apoptotic cells were defined as double positive for both active caspase 3 and cleaved PARP.

**Cell viability assay by MTS.** The described DLBCL and multiple myeloma cell lines were plated in duplicate at a density of 50,000 cells per well in 96-well plates along with DMSO as negative control, or different concentrations of IRAK1/4 inhibitor (EMD Chemicals). Cell viability at 1, 2 and 3 days after drug treatment was assayed by adding 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium and an electron coupling reagent (phenazine methosulphate; Promega), incubated for 3 h and measured by the amount of 490 nm absorbance using a 96-well plate reader. The presented data were derived from 3 days of drug treatment. The assay was performed twice.

**MYD88 and IRAK1 signature analysis.** To generate a gene expression signature of MYD88 signalling in ABC DLBCL, the HBL1 cell line was transduced with retroviral vectors expressing either shMYD88-4 or shMYD88-7. After puromycin selection, shRNA expression was induced for 24 or 48 h and gene expression was measured relative to parallel uninduced cultures using Agilent 4×44K oligonucleotide microarrays. A set of 284 MYD88 target genes was selected as those that were downregulated by 0.4  $\log_2$  in  $\geq 3$  arrays. A signature of NF- $\kappa$ B signalling (NF- $\kappa$ B-10 signature; <http://lymphochip.nih.gov/signaturedb/>) in ABC DLBCL was generated by treating HBL1 cells with the I $\kappa$ B kinase- $\beta$  inhibitor MLN120B for 2 h, 3 h, 4 h, 6 h, 8 h, 12 h, 16 h and 24 h. Genes that were downregulated  $>0.4 \log_2$  in at least four arrays with a one-sided *t*-test  $<0.01$  were chosen. A signature of JAK signalling in ABC DLBCL (JAKUp-2 signature; <http://lymphochip.nih.gov/signaturedb/>) was generated by treating HBL1 cells with JAK inhibitor I (5  $\mu$ M; Calbiochem) for 2 h, 4 h, 6 h and 8 h. Genes were chosen that were decreased in expression by  $>0.4 \log_2$  at  $\geq 3$  time points. A signature of IFN signalling (IFN-3; <http://lymphochip.nih.gov/signaturedb/>) was curated as the union between three published gene expression signatures of type I interferon signalling (IFN-1, IRF3-1 and Module-3.1 signatures; <http://lymphochip.nih.gov/signaturedb/>). A Fisher's exact test was used to calculate the significance of the overlap between the MYD88 signature and the other signatures.

Similar methods were used to generate a signature of IRAK1 signalling in ABC DLBCL. Two ABC DLBCL cell lines, HBL1 and TMD8, were transduced with retroviruses expressing shIRAK1-3 or a control shRNA. After puromycin selection, shRNA expression was induced for 24 h or 48 h and RNA and relative gene expression in shIRAK1 and control shRNA-expressing cells was analysed by gene expression profiling as above. A signature of 350 genes was selected as those that were downregulated by 0.4  $\log_2$  in  $\geq 3$  arrays.

26. Schmidlin, H., Diehl, S. A. & Blom, B. New insights into the regulation of human B-cell differentiation. *Trends Immunol.* **30**, 277–285 (2009).
27. Schmitz, R. *et al.* TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. *J. Exp. Med.* **206**, 981–989 (2009).

# Cell-type-specific replication initiation programs set fragility of the *FRA3B* fragile site

Anne Letessier<sup>1,2,3</sup>, Gaël A. Millot<sup>1,2,3</sup>, Stéphane Koundrioukoff<sup>1,2,3\*</sup>, Anne-Marie Lachagès<sup>1,2,3\*</sup>, Nicolas Vogt<sup>1,2,3\*</sup>, R. Scott Hansen<sup>4</sup>, Bernard Malfoy<sup>1,2,3</sup>, Olivier Brison<sup>1,2,3</sup> & Michelle Debatisse<sup>1,2,3</sup>

Common fragile sites have long been identified by cytogeneticists as chromosomal regions prone to breakage upon replication stress<sup>1</sup>. They are increasingly recognized to be preferential targets for oncogene-induced DNA damage in pre-neoplastic lesions<sup>2</sup> and hotspots for chromosomal rearrangements in various cancers<sup>3</sup>. Common fragile site instability was attributed to the fact that they contain sequences prone to form secondary structures that may impair replication fork movement, possibly leading to fork collapse resulting in DNA breaks<sup>4</sup>. Here we show, in contrast to this view, that the fragility of *FRA3B*—the most active common fragile site in human lymphocytes—does not rely on fork slowing or stalling but on a paucity of initiation events. Indeed, in lymphoblastoid cells, but not in fibroblasts, initiation events are excluded from a *FRA3B* core extending approximately 700 kilobases, which forces forks coming from flanking regions to cover long distances in order to complete replication. We also show that origins of the flanking regions fire in mid-S phase, leaving the site incompletely replicated upon fork slowing. Notably, *FRA3B* instability is specific to cells showing this particular initiation pattern. The fact that both origin setting<sup>5,6</sup> and replication timing are highly plastic<sup>7,8</sup> in mammalian cells explains the tissue specificity of common fragile site instability we observed. Thus, we propose that common fragile sites correspond to the latest initiation-poor regions to complete replication in a given cell type. For historical reasons, common fragile sites have been essentially mapped in lymphocytes<sup>1</sup>. Therefore, common fragile site contribution to chromosomal rearrangements in tumours should be reassessed after mapping fragile sites in the cell type from which each tumour originates.

Replication forks commonly face sequences intrinsically difficult to replicate such as natural pause sites<sup>9,10</sup>. Various observations have indicated that the sequence of common fragile sites may also constitute a challenge to fork movement. For example, it has been shown that replication is more delayed along common fragile sites than in the rest of the genome following treatment with aphidicolin<sup>11</sup>, a DNA polymerase inhibitor, and that the nucleotide sequences of common fragile sites contain subregions with the potential to form secondary structures<sup>4</sup>. Hence, it was suggested that helicases tend to travel uncoupled from polymerases in common fragile sites, giving rise to long stretches of single-stranded DNA (ssDNA) upon replication stress. In subregions able to adopt secondary structures, ssDNA formed ahead of polymerases may evolve into fork barriers that cause DNA breaks<sup>11</sup>. However, genome-wide analyses have not confirmed that common fragile sites are specifically enriched in sequences prone to form secondary structures<sup>12,13</sup>, and a recent study of the replication dynamics along *FRA6E* did not show fork slowing along the site<sup>14</sup>. The mechanism(s) involved in common fragile site instability thus have remained ill-defined.

We focused on *FRA3B*, the most active common fragile site in human lymphocytes. It overlaps the 1.5-Mb-long *FHIT* (fragile histidine triad)

tumour suppressor gene<sup>15,16</sup>, wherein lies the most fragile subregion of the site. To elucidate the molecular basis of *FRA3B* fragility, we analysed cells pulse-labelled successively with iododeoxyuridine (IdU) and then chlorodeoxyuridine (CldU), two thymidine analogues. We combined the DNA-combing technique with fluorescent detection of newly synthesized DNA and fluorescence *in situ* hybridization (FISH) with probes allowing identification of a 1.6-Mb-long region overlapping the *FHIT* gene (referred to as the *FHIT* locus) (see Methods, Supplementary Figs 1, 2 and Supplementary Table 1). This procedure permitted us to determine replication-fork speed, to search for potential fork stalling, and to map initiation and termination events along this large single-copy sequence (Supplementary Fig. 1).

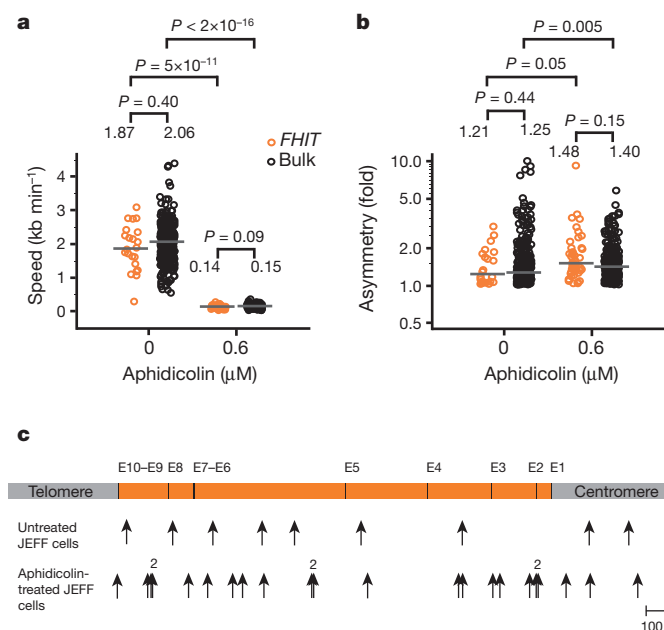
We studied Epstein–Barr-virus-immortalized human B lymphocytes (JEFF cells) (Supplementary Fig. 3), a cell type that shows a high frequency of breaks at *FRA3B* upon aphidicolin treatment (see later). We compared fork speeds along the *FHIT* locus and in the bulk genome in untreated and aphidicolin-treated cells (Fig. 1a). Statistical analyses show that differences found between treated and untreated cells are significant but not those between the locus and the bulk genome, regardless of the growth conditions. Fork speeds seem to be very heterogeneous along the locus, a phenomenon previously observed in a non-fragile region studied by molecular combing<sup>17</sup>. Plotting the speed of each fork as a function of its position along the locus shows that slow forks do not cluster in any particular subregion (Supplementary Fig. 4), which together indicates that the *FHIT* locus is not a mammalian equivalent of the slow-replicating zones described in yeast cells<sup>18</sup>.

Next we asked whether forks stall along the *FHIT* locus as they do at programmed pause sites in prokaryotes and yeasts<sup>19</sup>. Stalling should lead to asymmetrical forks, namely to individual forks presenting unequal IdU and CldU tracks (see Methods and Supplementary Fig. 1b). Therefore, we calculated asymmetry as the ratio of the longest to the shortest track (Fig. 1b). We found that fork asymmetry increases significantly upon aphidicolin treatment but, again, no differences were found between the locus and the bulk genome regardless of the growth conditions. Moreover, in untreated and aphidicolin-treated cells we found 9 and 22 forks along the locus, respectively, that have an asymmetry factor greater than or equal to 1.5. The mapping of these forks shows that they are evenly distributed (Fig. 1c). Lastly, mapping of all the forks—asymmetrical or not—travelling along the locus showed that no accumulation in particular subregions occurs in either untreated or aphidicolin-treated cells (Supplementary Fig. 5). Thus, forks do not specifically stall along the *FHIT* locus.

We then studied the distribution of initiation and termination events along the locus in untreated cells (Fig. 2a and Supplementary Fig. 1b). Notably, no initiation event maps to within an approximately 700-kb-long region (referred to as the *FRA3B* core) centred on exon 5. In contrast, ten and eight initiation events, respectively, take place in each of the approximately 500-kb-long regions flanking the core. Statistical analysis indicates that the deficit in initiation events in the core is

<sup>1</sup>Institut Curie, Centre de Recherche, 26 rue d'Ulm, 75248 Paris, France. <sup>2</sup>UPMC Univ. Paris 06, F-75005 Paris, France. <sup>3</sup>CNRS UMR 3244, F-75248 Paris, France. <sup>4</sup>Department of Medicine, Division of Medical Genetics, University of Washington School of Medicine, Seattle, Washington 98195, USA.

\* These authors contributed equally to this work.

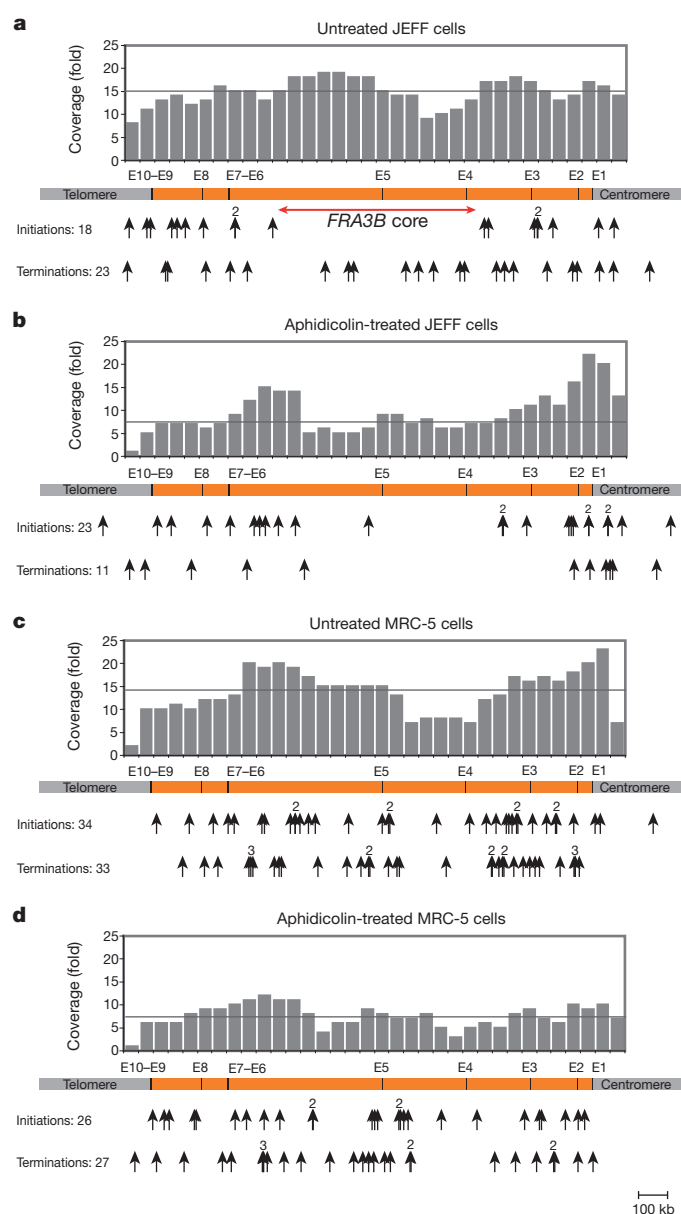


**Figure 1 | Comparison of fork properties in the *FHIT* locus and in the bulk genome in JEFF cells.** **a, b**, Distributions of fork speed (**a**) and fork asymmetry (**b**). Only non-truncated forks (see Supplementary Fig. 1) as determined by DNA counterstaining (see Supplementary Fig. 2) were recorded. Forks travelling in the locus are represented by orange circles (0 μM aphidicolin,  $n = 22$ ; 0.6 μM aphidicolin,  $n = 44$ ) and in the bulk genome by black circles (0 μM aphidicolin,  $n = 283$ ; 0.6 μM aphidicolin,  $n = 216$ ). Horizontal grey lines represent the medians of fork distributions. Medians and  $P$  values are indicated above the distributions. **c**, Upper panel shows *FHIT* gene (orange box) with its exons (E1 to E10). Arrows indicate position of forks with an asymmetry factor greater than or equal to 1.5 (0 μM aphidicolin,  $n = 9$ ; 0.6 μM aphidicolin,  $n = 22$ ). The number above some arrows indicates colocalized forks.

significant (see Methods). Large regions devoid of initiation events have been recently associated with transition zones separating domains with different replication timings. Long-travelling forks move unidirectionally along these zones, from the earliest to the latest domain, leading to progressive establishment of replication delays<sup>7,8,20</sup>. In contrast, forks travel in both directions within the *FRA3B* core (Supplementary Fig. 3) and termination events appear evenly distributed along the entire locus (Fig. 2a). Thus, in most cells, converging forks that emanate from each flanking region merge in the core, indicating that the *FHIT* locus is not a classical transition region.

We also mapped the initiation and termination events along the *FHIT* locus in aphidicolin-treated cells (Fig. 2b). We observed 23 initiation events, the distribution of which confirms the paucity of these events in the core. Despite a lower coverage of the locus the number of initiation events we observed in flanking regions seems comparable to that in untreated cells. Slowing replication speed thus triggers the recruitment of latent origins in flanking regions, as previously observed at other loci<sup>21,22</sup>. In addition, comparable levels of recruitment were observed in flanking regions and in the bulk genome (Supplementary Table 2). Notably, an approximately 900-kb-long region overlapping the core was devoid of any termination events. This deficit is not explained by preferential breakage of fibres bearing X-shaped structures in aphidicolin-treated cells because the density of termination remains normal in flanking regions. Instead, with complete replication of the core taking about 40 h with fork speed decreased to 0.14 kb min<sup>-1</sup>, terminations could occur only in cells that were close to completing replication when aphidicolin was added.

These results indicate that a paucity of initiation events contributes to fragility, which is in line with previous reports showing that the frequency of chromosome rearrangements increases in yeast mutants showing reduced replication-origin efficiencies<sup>23</sup>. In addition, a paucity



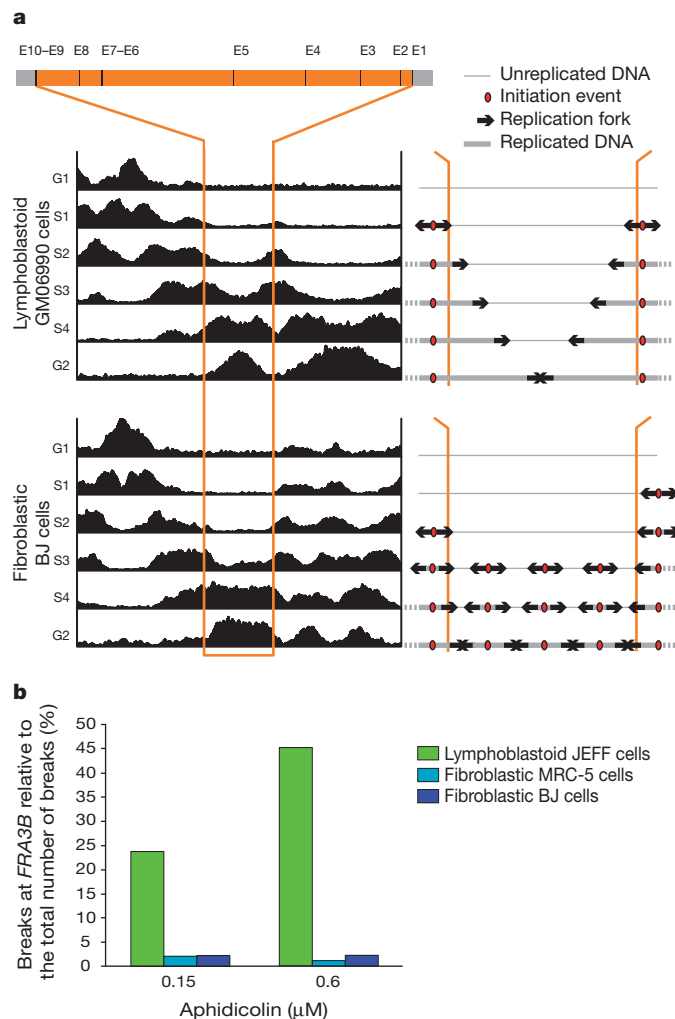
**Figure 2 | Mapping of initiation and termination events along the *FHIT* locus.** **a–d**, Untreated (**a**) and aphidicolin-treated (**b**) JEFF cells and untreated (**c**) and aphidicolin-treated (**d**) MRC-5 cells. Upper panels show histograms showing the coverage of the locus per 50-kb window. Each bar represents the number of fibres fully overlapping the corresponding sequence. Horizontal grey lines represent the medians of coverage. Middle panels are as in Fig. 1c. Lower panels, arrows indicate the position of initiation and termination events. The number above some arrows indicates colocalized events. The total number of events observed is indicated on the left.

of initiation events accounts for the *FRA3B* properties reported in the literature, as follows. Firstly, late replication completion, even in cells that go unperturbed through S phase<sup>24,25</sup>, may be due to the long distances that forks coming from flanking regions have to cover before merging. The abnormally long delay to complete replication upon aphidicolin treatment<sup>25</sup> is easily explained by the fact that slowing down fork movement affects long-travelling forks more profoundly than forks that need to cover only short distances. Secondly, increased instability upon inactivation of ATR, which stabilizes stalled forks, has been considered to support the fork barrier hypothesis<sup>11</sup>. Instead, polymerases travelling across long distances may need to be stabilized, notably upon replication stress. Lastly, reduced fragility in cells bearing deletions in the *FHIT* gene has been attributed to removal of sequences



prone to form secondary structures<sup>26</sup>. Alternatively, reducing the size of the core may alleviate incomplete replication in proportion to core shortening.

To characterize further *FRA3B* replication, we took advantage of a recent genome-wide analysis of replication timing by the Repli-Seq technique, which allows reconstitution of fork dynamics during the cell cycle<sup>7</sup>. Several human cell types were studied but information concerning common fragile sites was not exploited. Analysis of a 7-Mb-long region containing the *FHIT* locus (Fig. 3a and Supplementary Fig. 6) showed that the replication profiles along the locus differ notably between fibroblastic (BJ) and lymphoblastoid (GM06990 and H0287) cells. Whereas the profiles observed in lymphoblastoid cells are consistent with the results obtained in JEFF cells by combing, the plateau-shaped profile visible all along the *FHIT* locus in BJ cells indicates that initiation events take place within the *FRA3B* core in these cells. To



**Figure 3 | Relationship between replication profile of the *FHIT* locus and *FRA3B* fragility in lymphoblastoid and fibroblastic cells.** **a**, Left panel shows Repli-Seq analysis of the replication profiles along a 7-Mb-long region (chromosome 3: 56961124.5–63961124.5; assembly NCBI36 (hg18)) overlapping the *FHIT* gene (delimited by the orange lines) in lymphoblastoid GM06990 cells and fibroblastic BJ cells. Horizontal axis: position along the region. Vertical axis: density of sequence tags in each 50-kb window along the analysed region and for each indicated step of the cell cycle. Right panel shows cartoons illustrating the replication profiles along the *FHIT* locus (*FHIT* gene is delimited by orange lines). **b**, Number of breaks at *FRA3B* relative to the total number of breaks in cells treated with 0.15 μM or 0.6 μM of aphidicolin. Metaphases were prepared after a 3-h treatment with 200 nM nocodazole. See also Supplementary Table 3.

confirm and extend this conclusion, we used combing to study the replication profile of the *FHIT* locus in MRC-5 fibroblasts. Our results show that, both in untreated and aphidicolin-treated cells, initiation and termination events are evenly distributed all along the locus (Fig. 2c, d and Supplementary Fig. 7). Together, these observations are consistent with an increasing number of reports showing that both the replication timing<sup>7,8</sup> and the initiation pattern<sup>5,6</sup> of chromosomal domains evolve during differentiation but are conserved among different individuals in a given cell type<sup>27</sup>.

We reasoned that if the initiation-poor core of lymphoblastoid cells is involved in *FRA3B* instability, the site should not be fragile in fibroblasts. We used molecular cytogenetics to determine the frequency of breaks at *FRA3B* in aphidicolin-treated BJ, MRC-5 and JEFF cells (Fig. 3b, Supplementary Fig. 8 and Supplementary Table 3). Notably, the frequency of breaks at *FRA3B* is low in BJ and MRC-5 compared to JEFF cells. Long-travelling forks thus have a key role in *FRA3B* instability. However, this feature alone is insufficient to promote fragility as there are far fewer common fragile sites than initiation-poor regions in the genome. Moreover, it has long been shown that late replication completion contributes to *FRA3B* fragility<sup>25</sup>. Comparison of the replication timing in fibroblastic and lymphoblastoid cells (Fig. 3a and Supplementary Fig. 6) shows that replication of the *FHIT* gene ends very late in both cell types. Thus, late completion per se is also insufficient for fragility. These data indicate that common fragile site fragility results from the combination of late replication completion and paucity of initiation events. To test this prediction, we also analysed the Repli-Seq profiles along *FRA16D*, the second most active common fragile site in lymphocytes<sup>28</sup> (Supplementary Fig. 9a). We found that *FRA16D* shows tissue-specific initiation profiles and replication timing resembling those found at *FRA3B*. We also compared the frequency of breaks at *FRA16D* in aphidicolin-treated JEFF, MRC-5 and BJ cells (Supplementary Fig. 9b and Supplementary Table 4). Again the results look like those obtained for *FRA3B*, which reinforces and generalizes the mechanism we propose.

Tissue-specific reshuffling of the replication program provides a straightforward explanation for the setting of common fragile sites in different cell types, which would be very difficult to explain if nucleotide sequences were responsible for their fragility. In addition, the fact that cell-type-specific initiation profiles are maintained across species in regions of conserved synteny<sup>27</sup> agrees with the observed conservation of common fragile sites in different mammalian species<sup>11</sup>. Thus, we propose that common fragile sites are epigenetically defined loci that correspond to the latest initiation-poor regions to complete replication in a given cell type. The paucity of initiation events might reflect either a cell-type-specific lack of licensed origins<sup>29</sup> or chromatin organization delaying firing so much that cells enter mitosis before it occurs. As previously shown in yeast<sup>9</sup>, the mammalian checkpoint could be unable to delay mitotic onset when only a few long-travelling forks remain at work, which would allow unscheduled condensation of incompletely replicated common fragile sites, resulting in DNA breaks.

## METHODS SUMMARY

For combing experiments, JEFF or MRC-5 cells were grown for 5 h with or without 0.6 μM aphidicolin. Neo-synthesized DNA was labelled with a pulse of IdU followed by a pulse of CldU. After DNA combing, Morse-code probes were hybridized and fibres bearing replication and FISH signals were analysed. The replication fork speed variable was calculated using  $(d_i + d_{ci}) / (t_i + t_{ci})$  with  $d_i$  and  $t_i$  being the measured distance (in kb) and labelling time (in min) for IdU incorporation, respectively, and  $d_{ci}$  and  $t_{ci}$  the corresponding parameters for CldU incorporation. The fork asymmetry variable corresponded to  $\max(d_i / d_{ci}, d_{ci} / d_i)$ , which varies between 1 (no asymmetry) and  $+\infty$  (theoretical maximal asymmetry). Analysis of Repli-Seq data was carried out as described<sup>7</sup>. Breakage at common fragile sites was determined by FISH on metaphases with *FRA3B* or *FRA16D* bacterial artificial chromosome (BAC) probes. The total number of breaks was counted after Giemsa staining of metaphase spreads (see Methods).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 24 September; accepted 8 December 2010.**

**Published online 23 January 2011.**

1. Sutherland, G. R. & Richards, R. I. The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.* **5**, 323–327 (1995).
2. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability—an evolving hallmark of cancer. *Nature Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
3. Bignell, G. R. *et al.* Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
4. Schwartz, M., Zlotorynski, E. & Kerem, B. The molecular basis of common and rare fragile sites. *Cancer Lett.* **232**, 13–26 (2006).
5. Grégoire, D., Brodolin, K. & Méchali, M. HoxB domain induction silences DNA replication origins in the locus and specifies a single origin at its boundary. *EMBO Rep.* **7**, 812–816 (2006).
6. Dazy, S., Gandrillon, O., Hyrien, O. & Prioleau, M. N. Broadening of DNA replication origin usage during metazoan cell differentiation. *EMBO Rep.* **7**, 806–811 (2006).
7. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
8. Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).
9. Tourrière, H. & Pasero, P. Maintenance of fork integrity at damaged DNA and natural pause sites. *DNA Repair* **6**, 900–913 (2007).
10. Branzei, D. & Foiani, M. Maintaining genome stability at the replication fork. *Nature Rev. Mol. Cell Biol.* **11**, 208–219 (2010).
11. Durkin, S. G. & Glover, T. W. Chromosome fragile sites. *Annu. Rev. Genet.* **41**, 169–192 (2007).
12. Helmrich, A., Stout-Weider, K., Hermann, K., Schrock, E. & Heiden, T. Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Res.* **16**, 1222–1230 (2006).
13. Tsantoulis, P. K. *et al.* Oncogene-induced replication stress preferentially targets common fragile sites in preneoplastic lesions. A genome-wide study. *Oncogene* **27**, 3256–3264 (2008).
14. Palumbo, E., Matricardi, L., Tosoni, E., Bensimon, A. & Russo, A. Replication dynamics at common fragile site *FRA6E*. *Chromosoma*, (2010).
15. Huebner, K. & Croce, C. M. Cancer and the *FRA3B/FHIT* fragile locus: it's a HIT. *Br. J. Cancer* **88**, 1501–1506 (2003).
16. Pichiorri, F. *et al.* Molecular parameters of genome instability: roles of fragile genes at common fragile sites. *J. Cell. Biochem.* **104**, 1525–1533 (2008).
17. Lebofsky, R., Heilig, R., Sonnleitner, M., Weissenbach, J. & Bensimon, A. DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell* **17**, 5337–5345 (2006).
18. Cha, R. S. & Kleckner, N. ATR homolog Mec1 promotes fork progression, thus averting breaks in replication slow zones. *Science* **297**, 602–606 (2002).
19. Rothstein, R., Michel, B. & Gangloff, S. Replication fork pausing and recombination or “gimme a break”. *Genes Dev.* **14**, 1–10 (2000).
20. Farkash-Amar, S. *et al.* Global organization of replication time zones of the mouse genome. *Genome Res.* **18**, 1562–1570 (2008).
21. Anglana, M., Apiou, F., Bensimon, A. & Debatisse, M. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* **114**, 385–394 (2003).
22. Courbet, S. *et al.* Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature* **455**, 557–560 (2008).
23. Bielinsky, A. K. Replication origins: why do we need so many? *Cell Cycle* **2**, 307–309 (2003).
24. El Achkar, E., Gerbault-Seureau, M., Muleris, M., Dutrillaux, B. & Debatisse, M. Premature condensation induces breaks at the interface of early and late replicating chromosome bands bearing common fragile sites. *Proc. Natl Acad. Sci. USA* **102**, 18069–18074 (2005).
25. Le Beau, M. M. *et al.* Replication of a common fragile site, *FRA3B*, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum. Mol. Genet.* **7**, 755–761 (1998).
26. Durkin, S. G. *et al.* Replication stress induces tumor-like microdeletions in *FHIT/FRA3B*. *Proc. Natl Acad. Sci. USA* **105**, 246–251 (2008).
27. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
28. O'Keefe, L. V. & Richards, R. I. Common chromosomal fragile sites and cancer: focus on *FRA16D*. *Cancer Lett.* **232**, 37–47 (2006).
29. Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N. & Oda, M. Eukaryotic chromosome DNA replication: where, when, and how? *Annu. Rev. Biochem.* **79**, 89–130 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank E. Blackburn for critical reading of the manuscript. We thank Genomic Vision for making the DNA combing technology available to us. We acknowledge the Nikon Imaging Centre at Institut Curie-CNRS. We thank C. Rouzaud for help in combing experiments. A.L. is supported by a grant from the ARC (Association pour la recherche sur le cancer). The M.D. team is supported by La Ligue Nationale contre le Cancer (LNCC) (Equipe Labellisée EL2008.LNCC/MD), INCa (Institut National du Cancer) (2009-1-PLBIO-10-IC-1) and the PIC Réplication, Instabilité Chromosomique et Cancer (Institut Curie).

**Author Contributions** A.L. performed and analysed combing experiments. G.A.M. performed statistical and Repli-Seq analyses. S.K., A.-M.L. and O.B. performed cytogenetic analyses. N.V. and B.M. designed the Morse code. R.S.H. contributed to Repli-Seq analysis. A.L., G.A.M., O.B. and M.D. wrote the paper. M.D. planned the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.D. ([michelle.debatisse@curie.fr](mailto:michelle.debatisse@curie.fr)).

## METHODS

**Cell culture.** Epstein–Barr-virus-immortalized human B lymphocytes (JEFF cells) were grown in RPMI-1640 plus GlutaMAX-I medium (GIBCO). MRC-5 and BJ cells were grown in MEM plus Earle's salts without L-glutamine medium (GIBCO), 1% MEM nonessential amino acids (GIBCO), 1 mM sodium pyruvate (GIBCO) and 2 mM L-glutamine (GIBCO). In addition, all cells were grown with 10% fetal calf serum (PAN-Biotech GmbH), 100  $\mu\text{g ml}^{-1}$  of penicillin and streptomycin (GIBCO).

**FISH on metaphases and Giemsa staining.** JEFF or MRC-5 cells were grown for 16 h with aphidicolin 0.15  $\mu\text{M}$  or 0.6  $\mu\text{M}$  before metaphase preparations. Bacterial artificial chromosomes (BAC) were selected from the human March 2006 (NCBI36 (hg18)) assembly using the UCSC Genome Browser. *FRA3B* was probed with BAC RP11-32J15 (AC104161), RP11-641C17 (AC104164) and RP11-147N17 (AC104300). *FRA16D* was probed with BAC RP11-105F24 (AC106741) and RP11-571O6 (AC092724). Bacterial strains containing BAC were spread on LB agar plates containing 12.5  $\mu\text{g ml}^{-1}$  chloramphenicol and grown overnight at 37 °C. BAC DNA was then extracted according to the manufacturer's instructions (NucleoBond Xtra Midi Plus; Macherey–Nagel). Probes were biotinylated using the BioPrime DNA labelling system (Invitrogen) and purified on Illustra ProbeQuant G-50 Micro Columns (GE Healthcare). FISH on metaphases was performed as described<sup>24</sup> using 150 ng of each *FRA3B* BAC probe or *FRA16D* probe. Immunodetection was performed by successive incubations in the following reagents: (1) Alexa-488-conjugated streptavidin (Invitrogen); (2) biotin-conjugated rabbit anti-streptavidin (Rockland); (3) Alexa-488-conjugated streptavidin (Invitrogen). Chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI) (Vectashield mounting medium for fluorescence with DAPI; Vector Laboratories) and metaphases were observed by fluorescence microscopy. Total breaks, gaps or constrictions on chromosomes were observed on metaphase spreads stained by Giemsa (Giemsa R solution; Reactif RAL) without pre-treatment to obtain a homogeneous staining of chromosomes.

**DNA preparation for molecular combing.** JEFF or MRC-5 cells were grown for 5 h with or without aphidicolin 0.6  $\mu\text{M}$ . Neo-synthesized DNA was labelled as described<sup>21</sup> with the following changes: cells were pulse labelled for 30 min with IdU and for 30 min with CldU when grown in regular medium, or 2 h with IdU and 2 h with CldU when grown in the presence of aphidicolin. Genomic DNA was extracted and combing was performed as described<sup>30,31</sup>.

**FISH on combed DNA and immunofluorescence detection of neo-synthesized DNA.** Morse-code probes were designed as described<sup>17</sup>. Primer pairs and template BAC or fosmid used for PCR are listed in Supplementary Table 1. BAC and fosmid DNA were prepared as described earlier for FISH on metaphases. Hybridization was performed as described<sup>21</sup>. Immunodetection was performed by successive incubations in the following reagents: (1) Alexa-488-conjugated streptavidin (Invitrogen); (2) biotin-conjugated rabbit anti-streptavidin (Rockland); (3) Alexa-488-conjugated streptavidin (Invitrogen), mouse anti-bromodeoxyuridine (BrdU) (BD Biosciences) and rat anti-BrdU (AbD Serotec); (4) biotin-conjugated rabbit anti-streptavidin (Rockland), Alexa-350-conjugated goat anti-mouse (Invitrogen) and Texas-Red-conjugated donkey anti-rat (Jackson ImmunoResearch); (5) Alexa-488-conjugated streptavidin (Invitrogen), Alexa-350-conjugated donkey anti-goat (Invitrogen) and mouse anti-ssDNA (Millipore); (6) Cy5.5-conjugated goat anti-mouse (Abcam); (7) Cy5.5-conjugated donkey anti-goat (Abcam). Antibody incubations, washes and slide mounting were performed as reported previously<sup>21</sup>.

**Image acquisition and treatment.** An Eclipse 90i (Nikon) epifluorescence microscope connected to a Coolsnap HQ CCD camera and run by Metamorph software (Molecular Devices) was used for image acquisition. A  $\times 100$  objective was used for images of metaphases. A  $\times 60$  objective was used for combed DNA images of the *FHIT* locus and a  $\times 40$  objective for those of the bulk genome. For the locus, two overlays of images were set up for each microscope field. The first one combined the IdU/CldU and FISH signals to identify replicating fibres of interest. The second overlay combined the IdU/FISH and DNA signals to determine the length of the DNA fibre bearing the Morse code. Image analyses were performed with Photoshop and Illustrator (Adobe). Fork positioning along the locus corresponds to the barycentre of the IdU and CldU tracks:  $(d_1 + d_{cl})/2$ , with  $d_1$  and  $d_{cl}$  being the lengths (in kb) of the two tracks of the same fork.

**Repli-Seq analysis.** Analysis of Repli-Seq data was carried out as described<sup>7</sup>.

**Statistical analysis.** The R environment was used for all analyses<sup>32</sup>. Tables, graphical analyses and R code are available on request. Statistical significances were set to  $P \leq 0.05$  and  $\beta \leq 0.1$ . Two-tailed tests were systematically used. Type I and II errors were not controlled by any procedure of correction. The replication fork speed variable was calculated using  $(d_1 + d_{cl})/(t_1 + t_{cl})$ , with  $d_1$  and  $t_1$  being the measured distance (in kb) and labelling time (in min) for IdU incorporation, respectively, and  $d_{cl}$  and  $t_{cl}$  the corresponding parameters for CldU incorporation. The fork asymmetry variable corresponded to  $\max(d_1/d_{cl}, d_{cl}/d_1)$ , which varies between 1 (no asymmetry) and  $+\infty$  (theoretical maximal asymmetry).

Distribution of these two variables in each class was examined using histogram and quantile-quantile plot. In Fig. 1, data were compatible with the Mann–Whitney–Wilcoxon test. Normal approximation due to identical values and continuity correction were applied.

In Fig. 1, for non-significant  $P$  values, the 10% probability of being wrong when considering that the medians are not different is related to: (1)  $\Delta = 0.5 \text{ kb min}^{-1}$  without treatment and  $\Delta = 0.03 \text{ kb min}^{-1}$  in the presence of aphidicolin (Fig. 1a); and (2)  $\Delta = 0.62$  without treatment and  $\Delta = 0.80$  in the presence of aphidicolin (Fig. 1b).  $\Delta$  corresponds to the central parameter difference in the statistical populations verifying  $\beta \leq 0.1$  ( $\mu_{\text{Genome}} = \mu_{\text{FRA3B/FHIT}} + \Delta$ ,  $\mu$  being the central parameter in the population). It gives an estimation of confidence when accepting the  $H_0$  null hypothesis that there is no difference between the two central parameters (means/medians). Let us consider as an example the replication-fork speed in the absence of aphidicolin treatment: if the true difference between the *FRA3B/FHIT* locus median and the genome median is a minimum of  $0.5 \text{ kb min}^{-1}$ , then the probability of being wrong when accepting  $H_0$  is a maximum of 10%. In other words, a difference of  $0.5 \text{ kb min}^{-1}$  or less is considered to be non-relevant in this example when  $H_0$  is accepted. The procedure for getting  $\Delta$  for the replication-fork-speed variable between the two '*FRA3B/FHIT*' and 'genome' classes in the absence of aphidicolin is detailed as follows: the two distributions observed were approximately Gaussian, and thus were supposed to be the true distributions in the two statistical populations, except that the mean for 'genome' was set to  $\mu_{\text{FRA3B/FHIT}} + \Delta$ . Using R, two random samples of  $n_{\text{FRA3B/FHIT}} = 22$  and  $n_{\text{Genome}} = 283$  were generated, and the  $W$  statistic calculated. This procedure was repeated 5,000 times to get a  $W$  distribution. A  $\beta$  value was then obtained by determining the proportion of calculated  $W$  inside the  $(W_{0.025}, W_{0.975})$  interval, set by the  $W$  distribution when  $H_0$  is true ( $\Delta = 0$ ). This method was applied for increasing values of  $\Delta$  starting from zero. The whole procedure was repeated ten times and means of  $\beta$  were calculated for each  $\Delta$ . In all cases, the variation coefficient never exceeded 10%. Lastly, the  $\Delta$  minimal value that verifies  $\beta \leq 0.1$  was kept. The same procedure was applied for the fork asymmetry variable, except that observed distributions were approximately exponential (the fitted probability density was  $1/(m-1) \times e^{-(x-1)/(m-1)}$ ,  $m$  being the mean of the observed distribution). Note that for replication-fork-speed data, a Welch test was also used (distributions approximately normal) instead of the Mann–Whitney–Wilcoxon test. However, there was no consequence on the  $P$ -value result because of sample sizes.

The procedure we used to calculate that the deficit in initiation events in the core is significant both in untreated and aphidicolin-treated JEFF cells (Fig. 2a, b) is detailed as follows in the case of the experiment with untreated cells. Let us consider that the region without initiation events extends from the b5 to the d2 probe. The probability for one random initiation event occurring within the b5–d2 region could be defined as the size of this region divided by the size of the *FHIT* gene (from the beginning of E1 to the end of E10), which corresponds to  $674,643 \text{ bp} / 1,502,095 \text{ bp} = 0.45$ . However, this calculation does not take into account the coverage of the b5–d2 region and of the rest of the gene by the number of fibres analysed. Therefore, another procedure was chosen to calculate this probability. First, the size of the DNA segment harbouring an initiation event was defined as the distance in kilobases between the two divergent forks (fork sizes included). Thus, 13 measurements corresponding to the 13 initiation events in the *FHIT* gene (from the beginning of E1 to the end of E10) were obtained (range 13–526 kb). The lower value of this range gives an estimation of the minimal size of a DNA segment likely to contain a detectable initiation event. As the 13-kb value is considered low, the first quartile (102 kb) was taken as a threshold value below which a fibre is unlikely to carry such an event. Next, for each fibre, the part overlapping the b5–d2 region was measured and all the values over the 102-kb threshold were added (10,705 kb). The same was done for each fibre covering the rest of the gene (E1–b5 or d2–E10 areas, giving 12,555 kb). Thus, the probability of one random initiation event occurring in the *FHIT* gene localized inside the b5–d2 region was defined as  $10,705 / (10,705 + 12,555) = 0.46$ . Last, the probability of getting  $n$  events or less inside the b5–d2 region when 13 events occur randomly in the *FHIT* gene follows the binomial law  $B(13, 0.46)$ . For zero events,  $P(X=0) = 3 \times 10^{-4}$ . The same procedure was applied for aphidicolin-treated cells. The threshold obtained was 17.6 kb, the binomial law was  $B(18, 0.36)$  and  $P(X \leq 2) = 0.02$ .

30. Michalet, X. *et al.* Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* **277**, 1518–1523 (1997).

31. Labit, H. *et al.* A simple and optimized method of producing silanized surfaces for FISH and replication mapping on combed DNA fibers. *Biotechniques* **45**, 649–658 (2008).

32. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2006).



# MMSET regulates histone H4K20 methylation and 53BP1 accumulation at DNA damage sites

Huadong Pei<sup>1</sup>, Lindsey Zhang<sup>2\*</sup>, Kuntian Luo<sup>1\*</sup>, Yuxin Qin<sup>3</sup>, Marta Chesi<sup>4</sup>, Frances Fei<sup>2</sup>, P. Leif Bergsagel<sup>4</sup>, Liewei Wang<sup>3</sup>, Zhongsheng You<sup>2</sup> & Zhenkun Lou<sup>1</sup>

p53-binding protein 1 (53BP1) is known to be an important mediator of the DNA damage response<sup>1</sup>, with dimethylation of histone H4 lysine 20 (H4K20me2) critical to the recruitment of 53BP1 to double-strand breaks (DSBs)<sup>2,3</sup>. However, it is not clear how 53BP1 is specifically targeted to the sites of DNA damage, as the overall level of H4K20me2 does not seem to increase following DNA damage. It has been proposed that DNA breaks may cause exposure of methylated H4K20 previously buried within the chromosome; however, experimental evidence for such a model is lacking. Here we found that H4K20 methylation actually increases locally upon the induction of DSBs and that methylation of H4K20 at DSBs is mediated by the histone methyltransferase MMSET (also known as NSD2 or WHSC1) in mammals. Downregulation of MMSET significantly decreases H4K20 methylation at DSBs and the subsequent accumulation of 53BP1. Furthermore, we found that the recruitment of MMSET to DSBs requires the  $\gamma$ H2AX-MDC1 pathway; specifically, the interaction between the MDC1 BRCT domain and phosphorylated Ser 102 of MMSET. Thus, we propose that a pathway involving  $\gamma$ H2AX-MDC1-MMSET regulates the induction of H4K20 methylation on histones around DSBs, which, in turn, facilitates 53BP1 recruitment.

In response to DNA damage, 53BP1 rapidly relocates to the sites of DNA lesions in a phospho-H2AX ( $\gamma$ H2AX)- and MDC1-dependent manner<sup>4–7</sup>. 53BP1 is also recruited to the sites of DNA damage through a second mechanism that involves the binding of the tandem tudor domains of 53BP1 to methylated histones, with dimethylated H4K20 (H4K20me2) being the known physiological binding site for both mammalian 53BP1 and its yeast homologue Crb2<sup>2,3</sup>. However, unlike H2AX phosphorylation, no increase in the total levels of H4K20me2 was observed after DNA damage<sup>2,3</sup>. It is also not clear whether 53BP1 damage recruitment regulated by H2AX phosphorylation and H4K20 methylation are separate pathways or if they are interconnected. Studies from *Schizosaccharomyces pombe* showed that disruption of both H4K20 methylation and H2AX phosphorylation does not cause synergistic or additive effects on the DNA damage response, indicating that they might function in the same pathway<sup>8</sup>.

We examined H4K20 methylation at the sites of DNA damage using the cellular system (a HeLa clone carrying the DR-GFP homologous recombination reporter), in which expression of exogenous I-SceI introduces a single DSB in the cell's genome<sup>9</sup>. After I-SceI induction of DSBs, chromatin was immunoprecipitated from the cells using antibodies directed against mono-, di- or trimethylated H4K20 (H4K20me1/2/3), and quantitative polymerase chain reaction (qPCR) was used to determine the relative abundance of H4K20me1/2/3 at the induced break sites, while standard PCR gave a visual representation of the relative accumulation of these proteins at the DSB sites. Interestingly, H4K20me1/2/3 at the I-SceI break site all increased after DSB induction (Fig. 1a, b), as did the H4K20me2 signal at the sites of DNA damage induced by laser irradiation (Fig. 1c). Consistent with previous reports,

we did not observe apparent increase in total H4K20me2 levels<sup>2,3</sup> by western blot at commonly used ionizing radiation doses (Supplementary Fig. 1a), but we did observe a notable increase following high doses of ionizing radiation. This indicates that local increases of H4K20me2 at DSBs induced by low doses of ionizing radiation are masked from detection by western blotting owing to the high basal levels of H4K20me2 occurring throughout the genome.

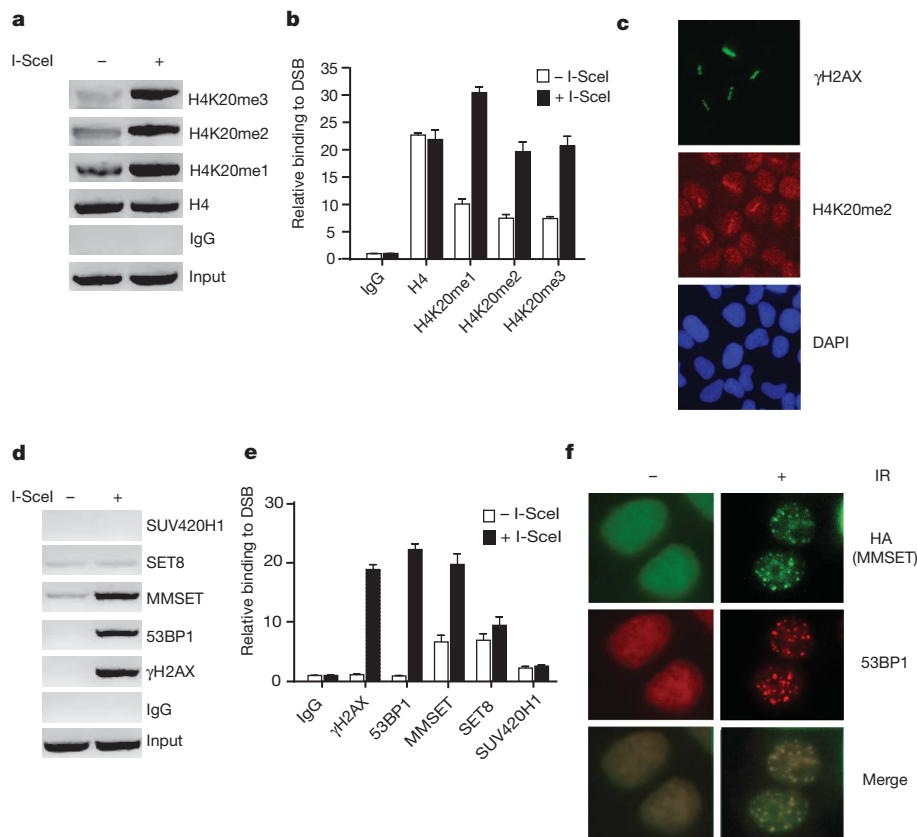
Next we investigated how the increase of H4K20me2/3 was induced at DSBs. It has been proposed that SET8 is mainly responsible for H4K20me1<sup>10,11</sup>, which is required for subsequent di- and trimethylation of H4K20. SUV420H1 and SUV420H2 are the major enzymes responsible for H4K20me2 and H4K20me3, respectively<sup>12,13</sup>. However, despite SUV420H1/2 loss and the subsequent lack of most H4K20me2/3, 53BP1 accumulation at DSBs was not abolished and only slightly delayed<sup>13</sup>. We did not observe substantial accumulation of SUV420H1 at the DSBs, whereas small amounts of SET8 localized to the I-SceI site both before and after DNA cleavage (Fig. 1d, e). This indicates that other histone methyltransferases methylate H4K20 specifically at DSBs. Interestingly, we found that MMSET, a newly identified histone methyltransferase<sup>14–16</sup>, accumulated at DSBs (Fig. 1d, e and Supplementary Fig. 1b).

Consistent with the results obtained from chromatin immunoprecipitation (ChIP) assays, MMSET formed discrete foci after ionizing radiation, colocalizing with 53BP1 (Fig. 1f). MMSET has been shown to methylate H3K36, H3K27 and H4K20<sup>14–16</sup>, and misregulation of MMSET due to haploinsufficiency in Wolf-Hirschhorn syndrome<sup>17</sup> and by t(4;14) chromosome translocation in multiple myeloma<sup>18,19</sup> indicates that it has an important role in the pathogenesis of these diseases. However, the cellular function of MMSET is largely uncharacterized. Our results imply that MMSET has a role in the DNA damage response (Fig. 1d–f). In support of this, downregulation of MMSET resulted in cellular hypersensitivity to ionizing radiation (Supplementary Fig. 1c).

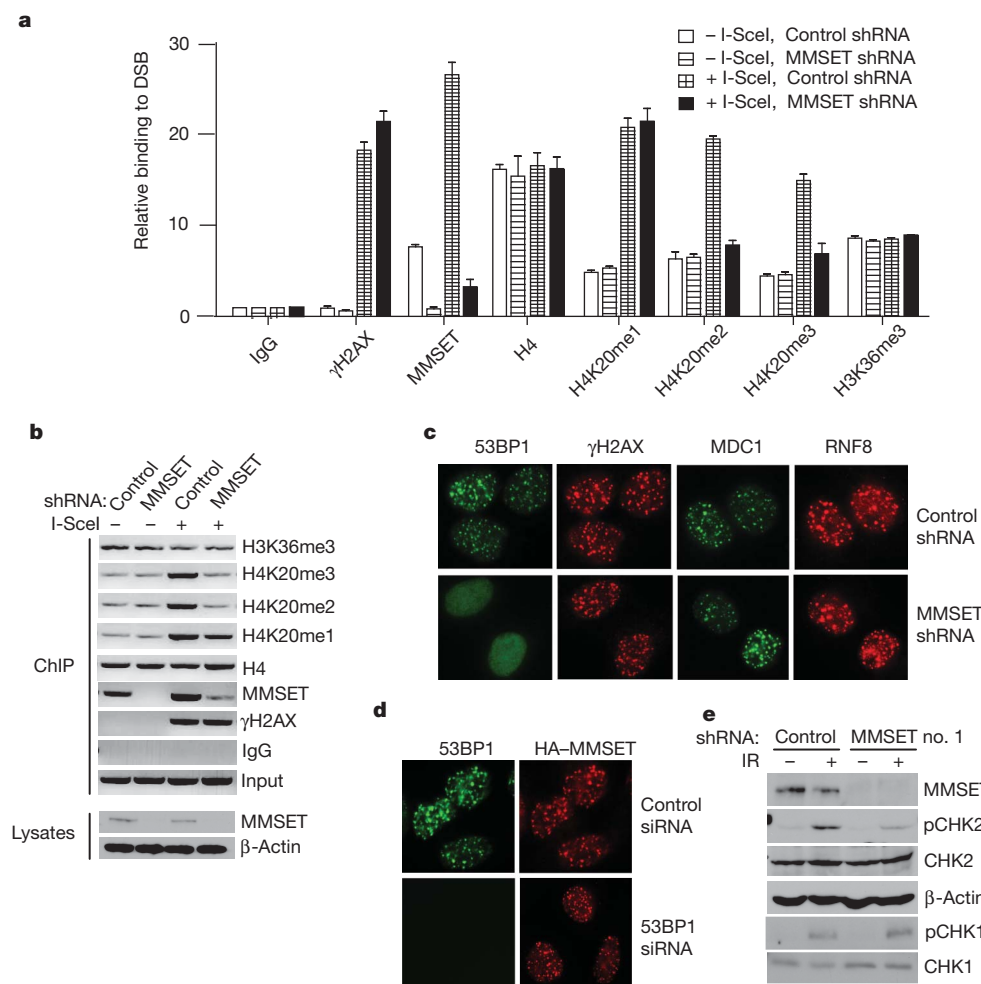
We proposed that MMSET regulates the DNA damage response through H4K20 methylation at DSBs. As shown in Fig. 2a, b and Supplementary Fig. 2a, b, downregulation of MMSET significantly decreased H4K20me2/3 at DSBs, but did not significantly affect H4K20me1 or H3K36 methylation at DSBs. We reasoned that if MMSET regulates H4K20me2 at DSBs, then MMSET should regulate the recruitment of 53BP1 to DSBs. Indeed, we found that downregulation of MMSET significantly decreased DNA-damage-induced focus formation of 53BP1, but not  $\gamma$ H2AX, MDC1 or RNF8, which are upstream regulators of 53BP1 (Fig. 2c and Supplementary Fig. 2c, d). Further, 53BP1 focus formation was defective in cells overexpressing a truncated MMSET (H929)<sup>15</sup>, whereas in cells expressing full-length MMSET (KMS11), 53BP1 focus formation was unaffected (Supplementary Fig. 2e). Downregulation of 53BP1 did not affect MMSET focus formation (Fig. 2d), indicating that MMSET is an upstream regulator of 53BP1. Importantly, downstream signalling events regulated by 53BP1, such as CHK2 phosphorylation<sup>20</sup>, were

<sup>1</sup>Division of Oncology Research, Mayo Clinic, Rochester, Minnesota 55905, USA. <sup>2</sup>Department of Cell Biology and Physiology, Washington University, St Louis, Missouri 63110, USA. <sup>3</sup>Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, Minnesota 55905, USA. <sup>4</sup>Comprehensive Cancer Center, Mayo Clinic Arizona, Scottsdale, Arizona 85259, USA.

\*These authors contributed equally to this work.



**Figure 1 | Induction of H4K20 methylation and recruitment of MMSET at DSBs.** **a, d,** Examples of ChIP analysis by PCR of indicated proteins on a DSB induced by I-SceI, where input demonstrates equal amount of DNA for ChIP. **b, e,** qPCR of indicated ChIP samples, where the y-axis represents the relative enrichment of the indicated proteins compared to the IgG control (after normalization with a PCR internal control to a locus other than the DSB; data  $\pm$  s.e.m.;  $n = 3$ ). **c, f,** Immunofluorescence staining of U2OS cells (**c**) and 293T cells (**f**) after indicated treatments, then stained with indicated antibodies. HA, haemagglutinin; IR, ionizing radiation.



**Figure 2 | MMSET is required for H4K20me2/3 and 53BP1 accumulation at DSBs.** **a,** qPCR analysis of ChIP samples from HeLa DR-GFP cells transfected with control or MMSET shRNA, where the y-axis represents the relative enrichment of the indicated proteins compared to the IgG control (data  $\pm$  s.e.m.;  $n = 3$ ). **b,** Ethidium bromide staining of ChIP samples from **a** analysed by PCR, where input demonstrates equal loading of DNA for PCR. **c, d,** Immunofluorescence of HCT116 cells transfected with the indicated siRNA or shRNA, irradiated (5 Gy), and stained with indicated antibodies. **e,** Phosphorylation (p) of CHK1/2 in the cell lysates from **c** analysed by immunoblotting.

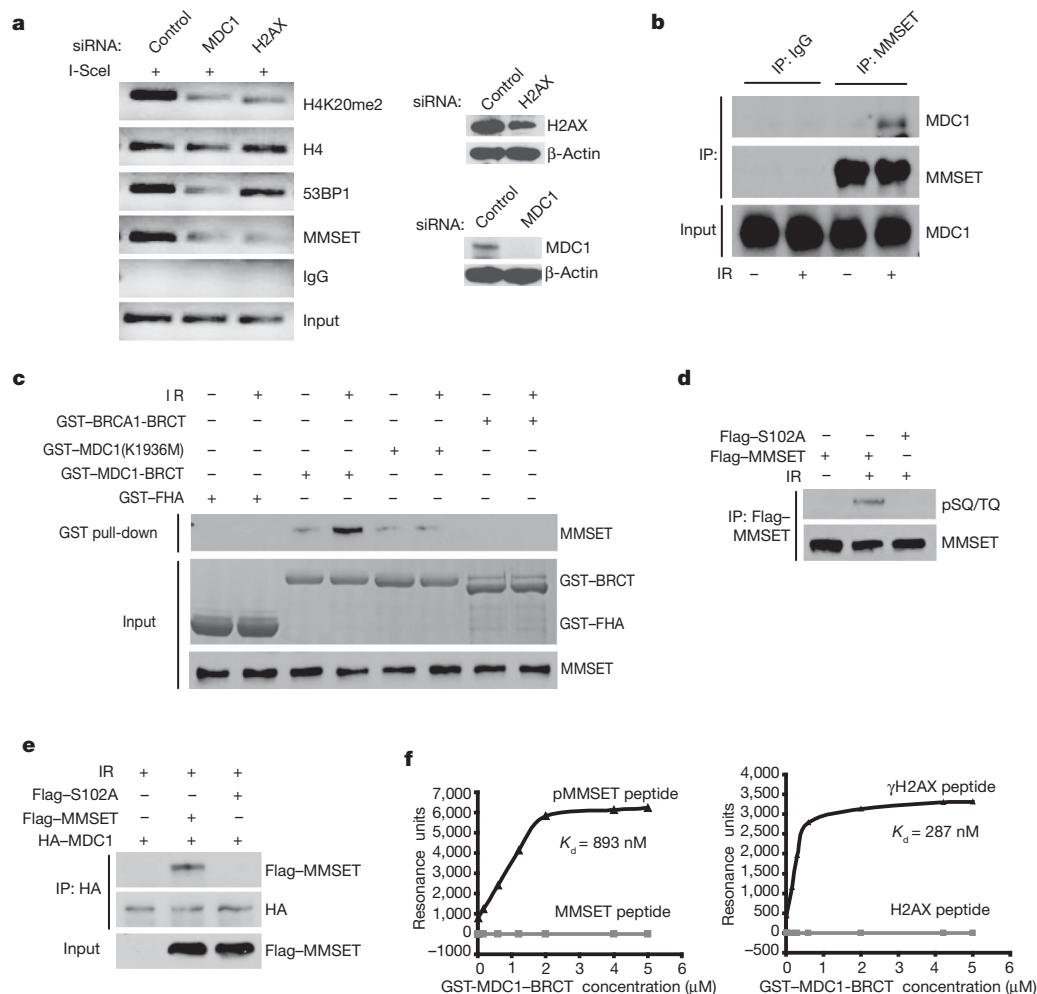
impaired by downregulation of MMSET (Fig. 2e and Supplementary Fig. 2f). To determine whether MMSET methyltransferase activity is required for these processes, we mutated the critical residue (F1117) required for MMSET methyltransferase activity<sup>15</sup>. We re-introduced short hairpin RNA (shRNA)-resistant wild-type MMSET or MMSET(F1117A) to cells stably transfected with MMSET shRNA. As shown in Supplementary Fig. 2g, h, whereas wild-type MMSET restored H4K20 methylation and 53BP1 recruitment to DSBs, MMSET(F1117A) did not. These data indicate that MMSET methylates H4K20 at DSBs, which facilitates the subsequent accumulation of 53BP1.

Previous studies indicated that the accumulation of 53BP1 at sites of DNA damage also requires H2AX and MDC1<sup>4–7</sup>. On investigation of this potential connection, we found that MMSET accumulation at DSBs was significantly reduced in cells depleted of H2AX and MDC1 (Fig. 3a and Supplementary Fig. 3a), as was H4K20me2 and the accumulation of 53BP1. Further, MDC1 foci appeared earlier than those of MMSET, whereas MMSET foci formed earlier than those of 53BP1 (Supplementary Fig. 3b). Thus, the accumulation of MMSET and the subsequent methylation of H4K20 and 53BP1 recruitment at DSBs seem to require H2AX and MDC1.

Previous studies also indicate that downstream of MDC1, the E3 ubiquitin ligase RNF8 regulates 53BP1 foci formation through its role in histone ubiquitination<sup>21–23</sup>. It is unclear whether RNF8 and MMSET

regulate 53BP1 accumulation in parallel or in the same pathway. As shown in Supplementary Fig. 4a, downregulation of RNF8 did not affect MMSET recruitment and H4K20 methylation, although 53BP1 recruitment was compromised. In addition, downregulation of MMSET had no effect on the recruitment of RNF8 to DSBs or the ubiquitination of H2A at DSBs (Fig. 2c, Supplementary Fig. 2c and 4b), indicating that RNF8 and MMSET function in distinct pathways. Thus, the mechanism through which RNF8-mediated ubiquitination events regulate 53BP1 recruitment remains to be determined.

While investigating how the H2AX–MDC1 pathway regulates MMSET accumulation at DSBs, we found that MMSET interacted with MDC1 in a DNA-damage-inducible manner (Fig. 3b). The interaction seemed to be specific to the MDC1 BRCT domain, as the BRCA1 BRCT domain and the MDC1 BRCT-domain mutant K1936M<sup>24</sup> were unable to interact (Fig. 3c and Supplementary Fig. 4c). Because BRCT domains recognize phospho-Ser/Thr motifs<sup>25,26</sup>, it is likely that MMSET is phosphorylated following DNA damage, thereby facilitating its interaction with MDC1. As shown in Supplementary Fig. 4d, MMSET was phosphorylated at ATM consensus SQ/TQ sites after ionizing radiation. No phospho-SQ/TQ signal was detected in ATM-deficient MEF cells or in samples treated with  $\lambda$ -phosphatase, indicating that MMSET is phosphorylated in an ATM-dependent manner. A previous large-scale proteomic study demonstrated that Ser 102 of MMSET is phosphorylated by ATM after



**Figure 3 | Recruitment of MMSET to DSBs requires the ATM–H2AX–MDC1 pathway.** **a**, ChIP analysis by PCR of indicated proteins at DSBs in HeLa DR–GFP cells transfected with the indicated siRNA. Right panels show western blots of H2AX and MDC1. **b**, Co-immunoprecipitation of MMSET and MDC1 in HeLa cells before or after ionizing radiation. **c**, GST pull-down assay of MMSET using indicated GST fusion proteins. **d**, 293T cells treated and

immunoprecipitated as indicated, then analysed with anti-pSQ/TQ antibody. **e**, 293T cells transfected with the indicated constructs were treated as indicated, then immunoprecipitated and immunoblotted with indicated antibodies. **f**, The interaction between GST–MDC1–BRCT and indicated peptides were measured by Biacore 3000.



DNA damage<sup>27</sup>. As shown in Fig. 3d, mutation at S102 abolished ATM-dependent MMSET phosphorylation after DNA damage, indicating that S102 is the major ATM phosphorylation site of MMSET. Further, mutation at S102 abolished the MDC1–MMSET interaction (Fig. 3e), verifying that the phosphorylation of S102 is required for MDC1 binding. The MDC1 BRCT domain has been shown to bind phospho-139 (pS<sub>139</sub>QEY) of  $\gamma$ H2AX, and a carboxy-terminal Y at the +3 position is critical for the binding specificity, although E at +2 is also positively selected<sup>24,28</sup>. The MMSET sequence after S102 is QEM, and does not contain Y at the +3 position. To confirm further the specificity of the MDC1–MMSET interaction, we used peptides containing either S102 or phospho-S102 to perform several assays. As shown in Supplementary Fig. 4e, phosphopeptides of MMSET preferentially pulled-down endogenous MDC1 from cell lysates. We determined further the binding affinity between MMSET peptides and the MDC1 BRCT domain using surface plasmon resonance (SPR). We found that the MDC1 BRCT domain preferentially bound MMSET phosphopeptides ( $K_d = 893$  nM), although with a lower affinity than it did  $\gamma$ H2AX peptides ( $K_d = 287$  nM). No MDC1 binding was found for non-phosphopeptides of MMSET (Fig. 3f).

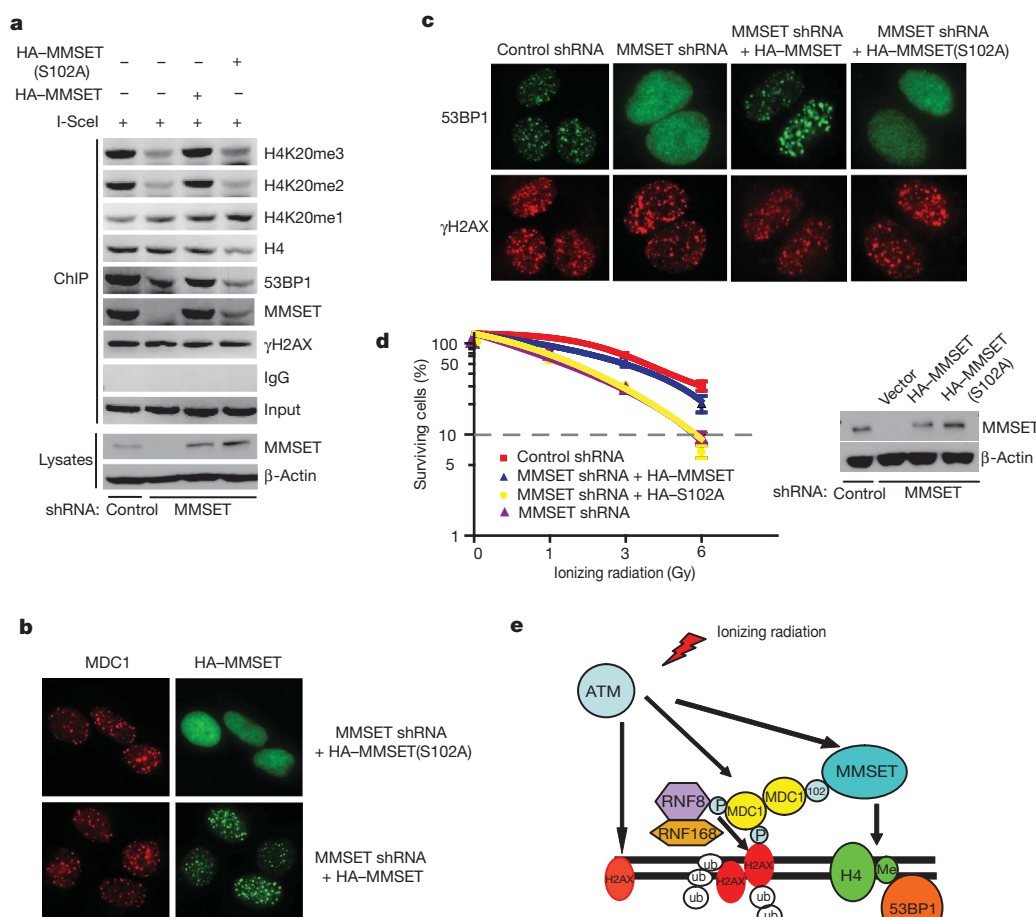
To investigate the functional significance of MMSET phosphorylation, we stably transfected HeLa DR–GFP cells with MMSET shRNA, and reconstituted these cells with shRNA-resistant wild-type MMSET or the MMSET(S102A) mutant. As shown in Fig. 4a, b and Supplementary Fig. 5a, wild-type MMSET was recruited to DSBs, but the recruitment of MMSET(S102A) was defective. This indicates

that S102 phosphorylation and the MDC1–MMSET interaction are essential for MMSET accumulation at DSBs. Similarly, reconstitution of wild-type MMSET, but not MMSET(S102A), rescued H4K20me2/3, 53BP1 accumulation at DSBs and CHK2 phosphorylation (Fig. 4a, c and Supplementary Fig. 5a–d). It is possible that the S102A mutation affects MMSET methyltransferase activity and subsequent H4K20 methylation. However, we found that the activity of MMSET and MMSET(S102A) towards histone H4 is comparable before and after ionizing radiation, indicating that the effects described earlier caused by S102A mutation are not due to a decreased methyltransferase activity (Supplementary Fig. 5e, f).

The BRCT domain of MDC1 is required for binding to  $\gamma$ H2AX at DNA damage sites, but it is unclear whether MDC1 uses this same domain to recruit MMSET to DSBs. We found that MDC1 formed oligomers (Supplementary Fig. 5g) and DNA damage increased the oligomerization of MDC1 (Supplementary Fig. 5h). Therefore, it is likely that different molecules in the MDC1 multimers bind  $\gamma$ H2AX and MMSET separately at the DNA damage sites.

Lastly, to examine how MMSET phosphorylation ultimately affects cellular sensitivity to DNA damage, we performed colony formation assays. Depletion of MMSET resulted in a significant increase in ionizing radiation sensitivity (Fig. 4d), and reconstitution with wild-type MMSET could reverse this effect whereas MMSET(S102A) could not.

Our studies reveal a critical role of the methyltransferase MMSET in regulating the assembly of 53BP1 foci at DNA lesions (Fig. 4e). We show that H4K20 methylation, unlike the previously held view, is



**Figure 4 | Phosphorylation of MMSET is important for H4K20 methylation, 53BP1 recruitment and the DNA damage response.** **a**, HeLa DR–GFP cells were transfected with the indicated constructs, H4K20 methylation and MMSET recruitment was analysed by PCR of ChIP samples. **b**, **c**, HCT116 cells transfected with indicated constructs were irradiated, and

10 min later, stained with indicated antibodies. **d**, Radiation sensitivity of cells from **c** was determined by colony formation (data  $\pm$  s.e.m.;  $n = 3$ ). **e**, Model demonstrating how the MDC1–MMSET pathway regulates DNA-damage-induced histone H4 Lysine 20 methylation and 53BP1 foci formation.

induced at DSBs. We also establish a previously unrecognized link between the H2AX-MDC1 pathway and H4K20 methylation, and show that MMSET connects these two pathways. These results indicate that multiple myeloma tumours with t(4;14) translocation and MMSET dysregulation may have aberrant responses to DNA damage, which may be related to the poor prognosis observed in this subgroup of patients that are treated with DNA alkylating agents.

## METHODS SUMMARY

HeLa DR-GFP and MDA-MB-231 ROS8 cell lines were used for the ChIP assays, which were subsequently analysed by PCR or qPCR. Co-immunoprecipitation was used to detect protein-protein interactions *in vivo* and SPR was used to detect the affinity for the protein and peptide interaction *in vitro*. Transient transfection of short interfering RNA (siRNA) or stable downregulation by shRNA was used to decrease the level of specific proteins. Immunofluorescence staining was used to visualize protein accumulation and localization after DNA damage.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 November 2009; accepted 9 November 2010.**

1. FitzGerald, J. E., Grenon, M. & Lowndes, N. F. 53BP1: function and mechanisms of focal recruitment. *Biochem. Soc. Trans.* **37**, 897–904 (2009).
2. Botuyan, M. V. *et al.* Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
3. Sanders, S. L. *et al.* Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell* **119**, 603–614 (2004).
4. Celeste, A. *et al.* Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nature Cell Biol.* **5**, 675–679 (2003).
5. Bekker-Jensen, S., Lukas, C., Melander, F., Bartek, J. & Lukas, J. Dynamic assembly and sustained retention of 53BP1 at the sites of DNA damage are controlled by Mdc1/NFBD1. *J. Cell Biol.* **170**, 201–211 (2005).
6. Lou, Z. *et al.* MDC1 maintains genomic stability by participating in the amplification of ATM-dependent DNA damage signals. *Mol. Cell* **21**, 187–200 (2006).
7. Stewart, G. S., Wang, B., Bignell, C. R., Taylor, A. M. & Elledge, S. J. MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature* **421**, 961–966 (2003).
8. Du, L. L., Nakamura, T. M. & Russell, P. Histone modification-dependent and -independent pathways for recruitment of checkpoint protein Crb2 to double-strand breaks. *Genes Dev.* **20**, 1583–1596 (2006).
9. Moynahan, M. E., Pierce, A. J. & Jasin, M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol. Cell* **7**, 263–272 (2001).
10. Fang, J. *et al.* Purification and functional characterization of SET8, a nucleosomal histone H4-lysine 20-specific methyltransferase. *Curr. Biol.* **12**, 1086–1099 (2002).
11. Nishioka, K. *et al.* PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9**, 1201–1213 (2002).
12. Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18**, 1251–1262 (2004).
13. Schotta, G. *et al.* A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes Dev.* **22**, 2048–2061 (2008).
14. Kim, J. Y. *et al.* Multiple-myeloma-related WHSC1/MMSET isoform RE-IIBP is a histone methyltransferase with transcriptional repression activity. *Mol. Cell. Biol.* **28**, 2023–2034 (2008).
15. Marango, J. *et al.* The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood* **111**, 3145–3154 (2008).
16. Nimura, K. *et al.* A histone H3 lysine 36 trimethyltransferase links Nkx2-5 to Wolf-Hirschhorn syndrome. *Nature* **460**, 287–291 (2009).
17. Bergemann, A. D., Cole, F. & Hirschhorn, K. The etiology of Wolf-Hirschhorn syndrome. *Trends Genet.* **21**, 188–195 (2005).
18. Chesi, M. *et al.* The t(4;14) translocation in myeloma dysregulates both *FGFR3* and a novel gene, *MMSET*, resulting in IgH/MMSET hybrid transcripts. *Blood* **92**, 3025–3034 (1998).
19. Stec, I. *et al.* *WHSC1*, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to *IgH* in t(4;14) multiple myeloma. *Hum. Mol. Genet.* **7**, 1071–1082 (1998).
20. Wang, B., Matsuoka, S., Carpenter, P. B. & Elledge, S. J. 53BP1, a mediator of the DNA damage checkpoint. *Science* **298**, 1435–1438 (2002).
21. Huen, M. S. *et al.* RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. *Cell* **131**, 901–914 (2007).
22. Kolas, N. K. *et al.* Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase. *Science* **318**, 1637–1640 (2007).
23. Mailand, N. *et al.* RNF8 ubiquitylates histones at DNA double-strand breaks and promotes assembly of repair proteins. *Cell* **131**, 887–900 (2007).
24. Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell* **123**, 1213–1226 (2005).
25. Yu, X., Chini, C. C., He, M., Mer, G. & Chen, J. The BRCT domain is a phospho-protein binding domain. *Science* **302**, 639–642 (2003).
26. Manke, I. A., Lowery, D. M., Nguyen, A. & Yaffe, M. B. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science* **302**, 636–639 (2003).
27. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
28. Lee, M. S., Edwards, R. A., Thede, G. L. & Glover, J. N. Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the  $\gamma$ -H2AX histone tail. *J. Biol. Chem.* **280**, 32053–32056 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Jasin for providing HeLa DR-GFP cells, and S. Baylin for providing MDA-MB-231 cells with inducible I-SceI expression. We thank M. Goldberg, X. Yu and M. Stucki for providing MDC1 and BRCA1 constructs. We thank M. Huen for providing 53BP1 expression constructs. We thank J. D. Licht for providing MMSET antibodies and B. Ho Park for providing MMSET shRNAs. This work was supported by the Richard Schulze Family Foundation and the NIH (CA130996 and CA151329; Z.L.), a grant from the American Cancer Society (IRG-58-010-52; Z.Y.) and a Siteman Career Award in Breast Cancer Research (Z.Y.).

**Author Contributions** H.P. designed and performed the experiments; L.Z., K.L., Y.Q. and F.F. performed some experiments; M.C. and P.L.B. provided essential tools; L.W., Z.Y. and Z.L. designed the experiments and supervised the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to Z.L. ([lou.zhenkun@mayo.edu](mailto:lou.zhenkun@mayo.edu)).

## METHODS

**Plasmids and shRNAs.** The MMSET(S102A) mutant was generated by PCR-based site-directed mutagenesis against full-length MMSET (pCEFL-MMSET-II). Wild-type MMSET or MMSET(S102A) was cloned into a pIRES2 vector containing S- and Flag-tag. shRNA-resistant constructs were made by introducing a silent mutation at the MMSET coding region (1666–1671; CTCGG to CTGCGA). The MDC1 FHA and BRCT domains were cloned into the pGEX4T-1 vector for bacterial expression of GST fusion proteins.

MMSET shRNA 1: 5'-GCACGCTACAACACCAAGTTT; MMSET shRNA 2: 5'-GCACAGTCTTCGGAAGAGAGACACAATCA; control shRNA: 5'-TTCAA TAAATCTTGGAGGT; MDC1 siRNA (MDC1 cDNA 58–76): UCCAGUGAA UCCUUGAGGUdTdT; control siRNA: UUCAUAAUUCUUGAGGUdTdT; H2AX siRNA: CAACAAGAAGACGCGAAUCdTdT; 53BP1 siRNA: 5'-AA GAUACUCCUUGCC UGAUAA-3'; RNF8 siRNA: 5'-AGAAUGAGCUCC AAUGUAUUU-3'.

**Antibodies and cell lines.** MMSET antibodies were provided by J. D. Licht or purchased from Abcam. Commercial antibodies used for ChIP were obtained from Upstate biotechnology ( $\gamma$ H2AX mouse monoclonal), Millipore (H4, H4K20me1/2/3, H2AUB), Active Motif (H4K20me2) and Novus (rabbit 53BP1). Antibodies against p53, pSQ/TQ, phospho-CHEK2, CHEK1 and phospho-CHEK1 were purchased from Cell Signaling. CHEK2 antibody was purchased from Millipore. RNF8 antibody was purchased from Abcam. MDC1 antibodies have been previously described<sup>6</sup>.

293T cells were cultured in RPMI 1640 medium supplemented with 10% fetal bovine serum (FBS). HCT116 and U2OS cells were cultured in DMEM supplemented with 10% FBS. HeLa DR-GFP cell lines were cultured in DMEM supplemented with 10% FBS and 2 ng  $\mu$ l<sup>-1</sup> puromycin. Mouse embryonic fibroblasts (MEFs) were cultured in DMEM containing 10% FBS and 5% ES.

**Immunoprecipitation, immunoblotting, and *in vitro* pull-down assays.** We prepared cell lysates, performed immunoprecipitation, and immunoblotting as previously described<sup>29</sup>. GST fusion proteins were bound to glutathione sepharose overnight at 4 °C. The beads were washed with PBS twice and incubated with cell lysates for 3 h at 4 °C. Beads were then washed with NETN buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40) three times, and proteins bound to beads were eluted by SDS sample buffer (100 °C for 12 min) and separated by SDS-PAGE for western blot analysis.

**Immunofluorescence staining.** Cells grown on coverslips were fixed with 3% paraformaldehyde solution in 1× PBS containing 50 mM sucrose at room temperature (22 °C) for 15 min. After permeabilization with 0.5% Triton X-100 buffer containing 20 mM HEPES pH 7.4, 50 mM NaCl, 3 mM MgCl<sub>2</sub> and 300 mM sucrose at room temperature for 5 min, cells were blocked with 5% goat serum for 1 h at room temperature, then incubated with primary antibodies at 37 °C for 20 min. After washing with PBS twice, cells were incubated with FITC or rhodamine-conjugated secondary antibodies at 37 °C for 20 min. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI). After a final wash with PBS, coverslips were mounted with glycerol containing paraphenylenediamine.

**ChIP.** Induction of a single DSB in HeLa DR-GFP cells was performed through transfection of the I-SceI expression plasmid. Twenty-four hours after transfection, about 5 × 10<sup>7</sup> cells were treated with 1% formaldehyde for 10 min at room temperature to crosslink proteins to DNA. Glycine (0.125 M) was added and incubated at room temperature for 5 min to stop the cross-linking. Cells were harvested and the pellets were resuspended in cell lysis buffer (5 mM PIPES (KOH), pH 8.0, 85 mM KCl, 0.5% NP-40) containing the following protease inhibitors: 1  $\mu$ g ml<sup>-1</sup> leupeptin, 1  $\mu$ g ml<sup>-1</sup> aprotinin and 1 mM PMSF; and incubated for 10 min on ice. Nuclei were pelleted by centrifugation (2,200g for 5 min). Nuclei were then resuspended in nuclear lysis buffer (50 mM Tris, pH 8.1, 10 mM EDTA, 1% SDS containing the same protease inhibitors as in cell lysis buffer) and sonicated to shear chromatin to an average size of 0.6 kb. Once centrifuged until clear, the lysates were precleared overnight with salmon sperm DNA/protein-A agarose slurry. Twenty per cent of each supernatant was used as input control and processed with the cross-linking reversal step. The rest of the supernatant (about 80% of the total) was incubated with 5  $\mu$ g of the indicated antibody overnight at 4 °C with rotation. Complexes were washed four times, once in high salt buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.1% SDS, 0.5% deoxycholate, 1% NP-40, 1 mM EDTA), once in LiCl buffer (50 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1% NP-40, 0.5% deoxycholate, 1 mM EDTA) and twice in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0). Beads were resuspended in TE containing 50 mg ml<sup>-1</sup> of RNase and incubated for 30 min. Beads were washed with water and elution buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>) was added for 15 min. Crosslinks were reversed by adding 10  $\mu$ g ml<sup>-1</sup> RNase and 5 M NaCl to a final concentration of 0.3 M to the elutants and incubated in a 65 °C water bath for 4–5 h. Two volumes of

100% ethanol were added to the precipitate overnight at –20 °C. DNA was pelleted and resuspended in 100  $\mu$ l of water, 2  $\mu$ l of 0.5 M EDTA, 4  $\mu$ l 1 M Tris, pH 6.5, and 1  $\mu$ l of 20 mg ml<sup>-1</sup> Proteinase K was added and incubated for 1–2 h at 45 °C. DNA was then purified and used in PCR reactions.

The PCR primers for ChIP, about 220 bp away from the I-SceI cut site, were as follows: forward, 5'-TACAGCTCCTGGGCAACGTG-3'; reverse, 5'-TCCTGCT CTTGGGCTTCTCG-3'.

Amplification was performed using the following program: 95 °C for 5 min, 1 cycle; 95 °C for 45 s, 56 °C for 30 s and 72 °C for 30 s, 30 cycles; 72 °C for 10 min, 1 cycle. A total of 12.5  $\mu$ l of the PCR products was applied to a 1.2% agarose gel and visualized by ethidium bromide staining.

**Quantitative analysis of ChIP samples.** qPCR was performed on a 7500 RT-PCR System (Applied Biosystems) using the SYBR Green detection system with the following program: 95 °C for 5 min, 1 cycle; 95 °C for 45 s and 62 °C for 45 s, 40 cycles. As an internal control for the normalization of the specific fragments amplified, a locus outside the region of the DSB was amplified, in this case *FKBP5*, using the input control sample as template. The internal control (*FKBP5*) primers were as follows: forward, 5'-CAGTCAAGCAATGGAAG AAG-3'; reverse, 5'-CCCGTGCCACCCCTCAGTGA-3'.

After qPCR amplification, the *FKBP5* input controls for untransfected (no DSB) and I-SceI transfected (DSB) were used to normalize the untransfected and transfected samples respectively. After normalization, the relative levels of the indicated proteins on a DSB were calculated by comparison of untransfected and I-SceI transfected samples to their respective IgG controls. All qPCR reactions were performed in triplicate, with the s.e.m. values calculated from at least three independent experiments.

**Biacore analysis.** Binding was analysed in a Biacore 3000 system. The relevant biotinylated peptides (MMSET peptide sequences: biotin-AKLRFESQEMKG; pMMSET peptide sequences: biotin-AKLRFEP(p)SQEMKG; H2AX peptide sequences: KKATQASQEQY;  $\gamma$ H2AX peptide sequences: biotin-KKATQApSQEQY) were bound to an SA sensor chip (GE Healthcare). The indicated concentrations of bacterially expressed GST-MDC1-BRCT in HBS-EP (HEPES-buffered saline with EDTA and polysorbate 20; 10 mM HEPES, pH 7.4, 0.15 M NaCl, 3 mM EDTA and 0.005% (v/v) polysorbate 20) were injected over the immobilized peptides at a flow rate of 80  $\mu$ l min<sup>-1</sup>. Interactions between each peptide and GST-MDC1-BRCT were analysed and steady-state binding was determined at each concentration. Regeneration of the sensor chip surface between each injection was performed with three consecutive 5- $\mu$ l injections of a solution containing 50 mM NaOH and 1 M NaCl.

***In vitro* histone methyltransferase assay.** HA-MMSET and HA-53BP1 were expressed and purified from 293T cells with haemagglutinin (HA) tag antibody and subsequent HA peptide elution. Recombinant histone 4 protein was from Upstate. *In vitro* histone methyltransferase assay was carried out according to the manufacturer's instructions (SAM510: SAM Methyltransferase Assay kit, G-Biosciences). In brief, all proteins were dialysed against 0.1 M Tris-HCl, pH 8.0, 20  $\mu$ M HA-MMSET (or HA-MMSET(S102A) mutant) and 20  $\mu$ M H4 (or HA-53BP1) was used for every reaction. Absorbances at 510 nm were measured every 10–30 s at 37 °C until the increasing absorbances reached a plateau or the reactions were stopped by boiling in SDS buffer, their contents separated by 15% SDS-PAGE, and the methylation of H4 was visualized by immunoblotting with anti-H4K20me2 antibodies (Upstate).

**Laser irradiation and immunofluorescence staining.** A partially customized 'laser-scissors' microirradiation system with an inverted microscope (Nikon, Ti-E), a laser ablation unit (Photonic Instruments, MicroPoint) and microscope automation and imaging software (Molecular Devices, MetaMorph) were used to introduce DNA damage in cultured cells. A 337-nm nitrogen laser (with 1–20 Hz repetition rate, 2–6 ns pulse duration and 120  $\mu$ J/pulse energy) transmits radiation through an optical fibre and a dye cell containing a solution that produces a 551-nm dye laser. The laser microbeam is then focused by a 63× (NA 1.4) oil immersion microscope objective. The total laser energy delivered to each focused spot was set by an attenuator plate (50% transmission) and the number of pulses. Cells were cultured on 35-mm glass-bottomed dishes (MatTek Cultureware, P35G-15-14-C) before laser irradiation.

Following laser irradiation, cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences) for 10 min at room temperature. Immunofluorescence staining was performed as previously described<sup>30</sup>. Cells were then imaged using the Nikon microscope and the MetaMorph software described above.

29. Kim, J. E., Chen, J. & Lou, Z. DBC1 is a negative regulator of SIRT1. *Nature* **451**, 583–586 (2008).

30. You, Z. et al. CtIP links DNA double-strand break sensing to resection. *Mol. Cell* **36**, 954–969 (2009).

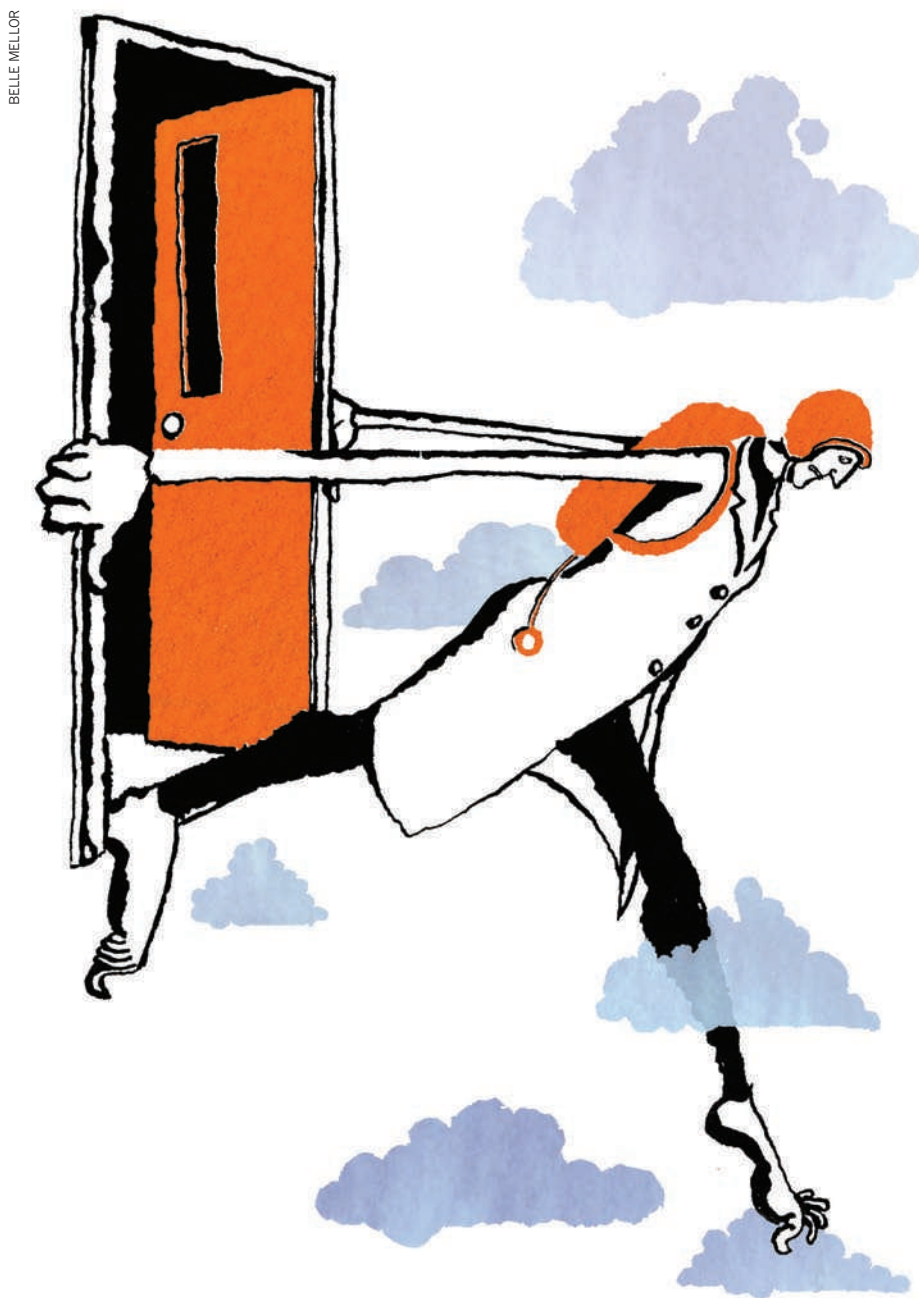


# CAREERS

**TURNING POINT** Inventing a sequencer was a milestone in a biotechnology career **p.131**

**GUIDANCE** Watch video seminars on the trends shaping bioscience [go.nature.com/aujxih](http://go.nature.com/aujxih)

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



BELLE MELLOR

BY LAURA BONETTA

**B**log author Biochembelle's post about her decision to leave a postdoc position struck a chord with many readers. She revealed an unsettling trend: a successful graduate student who had worked, published, defended her thesis and found the postdoc position of her dreams, only to discover some months later that those dreams had become a distant memory. The reality is that the 'sure-fire home-run' project isn't working out, the new department and environment aren't such good fits after all, and there are personality clashes with the principal investigator. "You are tired, angry, bitter, depressed," she wrote. "You have turned into the 'disgruntled doc' that you swore you'd never become."

Biochembelle, who asked to be referred to by the pseudonym she uses on her blog, has moved from one postdoc position in the United States to another, at a research and teaching hospital. She says she heard from many researchers in similar situations who, like her, chose to switch labs for various reasons. "For me, there were enough issues going on that I did not feel like I could do my best work in the environment I was in," says Biochembelle. "I don't think it is ever just one thing that pushes you to the point where it is time to walk away."

Regardless of the motivation, switching labs is not easy and postdocs should think carefully about it. Sometimes, problems can be resolved by a frank discussion with the principal investigator, perhaps by enlisting the help of a trusted mentor or adviser. But if walking away seems to be the best course of action, there are some steps that can help to ensure a smooth transition with few, if any, negative repercussions.

When informing their principal investigator of the move, postdocs should focus on career goals and opportunities rather than on personality issues. They should also give plenty of notice and find a way to either complete projects or leave them in such a state that they can be continued by someone else. And, of course, they should choose their next labs wisely. These measures can help postdocs to nourish their careers and stave off stagnation as they make their next moves.

## MULTIPLE POSITIONS

Many life scientists complete more than one postdoc before finding a permanent position. According to the US National Science Foundation, in 2006 about 43% of all doctorates in the biological, agricultural and environmental

## CAREER DECISIONS

# Taking the plunge

*Switching to a new postdoc may be risky and challenging, but it does not have to be career-threatening.*

► life sciences had held one postdoc position, 16% two postdoc positions, and 3.6% three or more. And the reasons are varied. “A lot of people do more than one postdoc not because the first postdoc did not work out but because they did not find a regular career position or because they wanted to get additional training,” says Thomas Gething, director of the office of postdoctoral affairs at the University of Washington Graduate School in Seattle. However, spending too much time as a postdoc can have negative consequences, especially in countries that have forced retirement at a certain age. In Germany, for example, the retirement age is 67, so time is precious. “If you work backwards, you need to be a full professor by your mid-40s,” says Jonathon Howard, director of the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany. “We tend to hire group leaders in their early 30s so if you are spending more than 5 years total as a postdoc I would think you would run into problems.”

This time constraint means that postdocs cannot afford to linger in unproductive positions. “Having to switch postdoc labs may not be an ideal situation, but sticking it out can be even worse,” says Rania Sanford, assistant dean for postdoctoral affairs at Stanford University in California. Staying in an unworkable postdoc to demonstrate commitment could be more detrimental than moving on, she says.

Howard agrees. After completing a PhD in Australia, he spent less than a year in his first postdoc in Britain before switching to a lab where he felt he could be more productive. “I knew after about six months that I should move on,” he says. The negative consequences were minimal, he says, because it was such a short stint. “I think it would be worse to stay five years in a postdoc and not get anything out of it.”



**“You want to make sure you don’t go from one bad situation to another.”**

Rania Sanford

## TOUGH CHOICES

But although switching labs can have career benefits, it is not always easy or practical to do. This is especially true for postdocs with visa concerns. Moving to a new university requires a lot of paperwork and there is no guarantee that a new visa will be granted in time. In extreme cases, leaving a lab might mean leaving the country. It can also run counter to a deeply ingrained cultural milieu. In China, “we have a deep belief that you have to face a problem head on and not give up,” explains Stanford University postdoc Xiaomeng Milton Yu. “If I were to say to my Chinese friends that I don’t get along with my supervisor and want to leave the lab, they may see that as giving up.” Although Yu is



Nick Brooks and Jonathon Howard suggest tackling issues with supervisors to avoid career problems.

happy in his current position, he says he knows of a few Chinese postdocs who have left their labs because they were unhappy with them. But, he says, “I think in general foreign postdocs are more likely to stick it out.”

Postdoc problems often arise because of differing expectations between the principal investigator and the postdoc related to project focus, productivity, research style and the postdoc’s career-development goals. A postdoc might lament the research time that they need to sacrifice to supervise others in the lab; the project might not fit the postdoc’s interests or career goals, or it might require skills that the postdoc lacks.

In many cases, such misunderstandings can be resolved simply by opening the lines of communication between principal investigator and postdoc. “Many times they are worried about the same thing,” says Jo Handelsman, a professor of molecular, cellular and developmental biology at Yale University in New Haven, Connecticut. Both might, for example, worry that the postdoc has not yet published any work. “The worst thing is leaving things to fester,” says Handelsman.

When direct communication isn’t possible, a postdoc might consider reaching out to other colleagues — perhaps a department chair, another principal investigator or staff in the postdoc or ombudsman office. At Imperial College London, all chemistry postdocs are offered an academic mentor, separate from their principal investigator, whom they get to know and can go to for “confidential advice of any sort,” says postdoc Nick Brooks, head of Imperial’s Chemistry Postdoc Development Team. “This system can help to mitigate potential problems between postdocs and principal investigators.”

But talking has its limits when stark personality clashes arise. “This is a much trickier problem to fix than a conflict in goals,” says Dorothy Shippen, a biochemist at Texas A&M University in College Station. In such cases, says Shippen, postdocs should think strategically about the decision to leave. First, she says, they should consider whether they can stay long enough to get something accomplished, such as a publication



HOWARD: MPI-CBG

or a good letter of recommendation. But if they are losing respect for their principal investigators, losing their love of science, or expecting to accomplish little by staying, they might need to look for new positions, she says.

A lack of funding can be an even more formidable obstacle to a successful postdoc experience than disagreements with the principal investigator. In some places, postdocs can be reliant on the principal investigator for grant money. According to the National Science Foundation, in the autumn of 2006, 56% of science and engineering postdocs at US universities were funded through federal research grants, up from 52% in 1993. Grant applications being denied can mean that a principal investigator cannot afford to keep all of his or her postdocs. And if a lab has an uncertain financial future, and the principal investigator is waiting for the results of various grant applications, a postdoc might find it prudent to seek a position that promises more grants. “That is happening a lot right now,” says Lynn Zechiedrich, a principal investigator and microbiologist at Baylor College of Medicine in Houston, Texas. But although a lack of funding can present an opportunity to let go of unproductive staff, principal investigators are often prepared to go the extra mile for those who are worth it. Zechiedrich, for instance, was ready to forgo her own pay rise to pay a postdoc’s salary while waiting for a grant to come through; another postdoc took a one-month furlough until the funding was available. “These were smart, hard-working postdocs, and now their results have helped us get more grants funded, so sometimes you have to get creative to maintain a postdoc position,” says Zechiedrich.

## EXIT PLAN

Whatever the reason, the decision to switch labs should not be a purely emotional reaction, says Sibby Anderson-Thompkins, director of the office of postdoctoral affairs at the University of North Carolina at Chapel Hill. “Make a list of pros and cons based on your expectations and career goals and on the overall

L. A. CICERO/STANFORD NEWS SERVICE



environment and relationships,” she says. “Then ask yourself if moving to another lab will give you more opportunities and more viable projects that result in papers or publications.”

As they weigh up their options, postdocs should consider talking discreetly to other principal investigators about joining their labs. It's also a good idea to learn as much as possible about the principal investigator and lab environment from current and past lab members. “You want to make sure you don't go from one bad situation to another,” says Sanford. She advises that postdocs find out exactly what would be expected of them. They should discuss the skills and training they have, and what they need to develop in the next year. “It is also important to understand what the lab direction is, what the principal investigator wants to do, and what the grants situation is,” she says, “so that there are no surprises.”

When approaching new labs, a postdoc should avoid disparaging the principal investigator of the lab he or she is leaving. “You can say there were challenges, but focus on the lessons learned and skills gained,” says Anderson-Thompkins. “It is okay to say that you wanted to pursue other opportunities, but you don't have to say how bad the lab or the principal investigator was.”

If, after careful consideration, a postdoc decides to leave his or her lab, the postdoc should inform the current principal investigator of the decision promptly, and make sure the conversation focuses on professional rather than personal issues.

Discussions should also focus on finishing existing projects or handing them over to other members of the lab. “Give them plenty of lead time and wrap up what is going on,” says Anderson-Thompkins. “That will help you leave on the best possible terms.” It could also mean that the postdoc is still able to garner a supportive letter of recommendation from the principal investigator in the future. And even if a glowing recommendation is out of the question, chances are that the ‘old’ principal investigator will be a collaborator or grant reviewer or a close friend to someone on a hiring committee. “You want to walk out the door with a good reputation,” says Anderson-Thompkins. “Don't do anything that will hurt your career.” ■

**Laura Bonetta** is a freelance writer based in Garrett Park, Maryland.

## TURNING POINT

# Jonathan Rothberg

*Last December, Jonathan Rothberg, founder and chief executive of Ion Torrent, a biotechnology company based in Guilford, Connecticut, released the Personal Genome Machine. The US\$50,000 desktop DNA sequencer will, he says, greatly improve access to genome sequencing.*

### What decision was pivotal in your early career?

I was interested in chemistry and engineering in high school, and did a chemical engineering undergraduate degree at Carnegie Mellon University in Pittsburgh, Pennsylvania. But my interests in biology and cognitive psychology were growing, and I had to decide which to follow for a PhD. I knew I wanted a set of tools that would make me marketable. The explosion in biology from genome sequencing set me up to combine my interests in computers, biology and engineering, and have an impact in a rapidly emerging field. So I got a PhD in biology from Yale University in New Haven, Connecticut.

### What is your advice to young scientists?

Master a number of fields. There will always be someone better than you at physics, maths or chemistry, but if you focus on mastering a few things you love, nobody will be better at that intersection.

### Who had the biggest influence on your career?

Steve Jobs [co-founder of Apple]. I loved the way he was changing the world in 1984. I saw him give a presentation in which he said the most profound thing I had heard — that the reason he had become influential was that he ‘just did it’. I know it sounds like a Nike commercial, but it hit home that most people simply think about things, and don't do them.

### Are you a scientist, inventor or entrepreneur?

I would say scientist and inventor. I am not an academic so I don't publish very often, but my publications have been on the covers of *Nature* and *Science*. I'm an entrepreneur only because assembling smart people and funding is essential to bringing inventions to market. But scientific needs inspire my inventions. For example, my newborn son had a health scare in 1999. The doctors had no way to tell whether he had an inherited disease, and I realized that an invention able to sequence an individual genome quickly would be useful. That idea sparked my second company, 454 Life Sciences. But my inventions also give me access to interesting, ground-breaking science. I cold-called Svante Pääbo, a geneticist at the Max Planck Institute for Evolutionary Anthropology in



Leipzig, Germany, and told him that I had a machine to help sequence the Neanderthal genome — which led to a collaboration.

### Is the Personal Genome Machine a turning point just for your career or for science in general?

I hope it is pivotal for science in general. We made a semiconductor device that sees chemistry in real time. A chip measures electrical charges during DNA replication, which lets it decode the sequence. It's a connection between chemistry and the digital world. This means that the sequencing machine will one day be as ubiquitous and cheap as the mobile phone.

### What skills do you think will be most in demand in the coming decade?

Quantitative skills — the ability to do calculations and estimations. Biology is great, but you need analytical skills. It no longer helps simply to describe something. We need more people at the intersections of fields. For example, bioinformaticians don't have to have a PhD in molecular biology, but they need enough of an understanding to develop an intuition about how systems work.

### How should would-be inventors go about bringing a technology to market?

They should do the hardest experiment, the one that poses the biggest obstacle to success, first — otherwise they could find themselves ten years later having made little progress. Many people lose themselves by not asking tough enough questions about their own inventions. If you can't clear the biggest hurdle, you are wasting everyone's time. ■

INTERVIEW BY VIRGINIA GEWIN



**“The worst thing is leaving things to fester.”**

Jo Handelsman



# TO THE STARS

Trade mission.

BY KEN LIU AND SHELLY LI

*Ad Astra Community Forum*

**twinkle\_twinkle (moderator):** Last week, ISS scientists announced that they've confirmed spectral signatures of water, methane, ozone and carbon dioxide in the atmosphere of the fourth planet orbiting Gliese 581. Hopefully, intelligent beings occupy this planet. As ISS scientists have not detected any electromagnetic radiation, we can assume that Gliese 581 is not any more technologically advanced than Earth. The natural next step is to establish contact. Marco Polo forged his way down the Silk Road in the late thirteenth century, fostering trade between Europe and Asia. Likewise, Ad Astra prides itself on its pioneering investment in space exploration. We have decided to fund a private trade mission to Gliese 581, to be launched next year. Ad Astra predicts much potential for interstellar trade in the years to come. Estimated flight time is 40 years, so the maximum payload we can transport is 200 kilograms and nonperishable (though if anyone knows of a way to get cheap antimatter so we can boost the payload, feel free to mention in the comments). Mostly importantly, Ad Astra is seeking your suggestions. What should we carry to Gliese 581? As long as it fits the requirements listed above, we will be happy to hear your thoughts. If we use your suggestion, you may receive an honorarium of one-tenth of 1% from our profits. (Note: the right to receive the honorarium is not assignable or transferable by inheritance.)

**nolo\_contendere (forum astronaut, second class):** A non-assignable royalty of 0.1%? Are you kidding? As it'll be at least 80 years before your ship even gets back, only kids will take up your offer. Good luck shipping pokemons and cotton candy to methane breathers.

**iheartlucy (forum space cadet):** Haven't they been awash in signals from our reruns for decades with nothing to watch them on? Can we ship them TV sets and charge for every show, INCLUDING COMMERCIALS? It'll be the ultimate syndication market.

**Anon\_4437 (forum guest):** Is everyone asleep here? The first priority ought to be to conquer them. If they haven't got to radio yet, we should send a platoon of Marines and take them down like Pizarro did in Peru. Then we can take whatever we want.

**iamnotneilarmstrong (forum astronaut, second class):** 4437, check your reading



comprehension. 200 kilos is the limit. And NON-PERISHABLE.

**Anon\_4437:** Oh, right. How about one fully equipped attack drone then?

**twinkle\_twinkle:** 4437, grow up. This is a trading mission of peace.

**Anon\_4437:** You don't get to the top of the food pyramid by being politically correct wimps. I'm giving you the most economically efficient suggestion.

**twinkle\_twinkle:** I've banned 4437's IP. Please keep the discussion focused.

**iheartlucy:** Does anybody know how to delete a previous post? I have a friend who might get me a meeting with some TV execs, and I wanna keep my idea for them. I copy-right it. I patent it. Don't take it.

**iamnotneilarmstrong:** No one cares. How do you know if the aliens even have eyes? Or a sense of humour?

**triune (forum astronaut, first class):** What about just trading for ideas? With the cost of fuel, it's uneconomical to trade even for diamonds. Probably best to use the weight allotment for a powerful transmitter, then we can just phone back and forth and share discoveries.

**iamnotneilarmstrong:** Long time no see, triune! Where've you been? I like your idea, not the least because it cuts down travel time to light speed, so only 40 years round-trip. What are the first ideas we want to share with them?

**triune:** Democracy, pacifism and a dose of the greatest hits from our religions.

**veryliberal (forum guest):** triune, how is this better than Anon\_4437's very sensible suggestion? You want to brainwash them with opiates for the masses and then take their stuff? At least Anon\_4437 was honest.

**twinkle\_twinkle:** I've

banned veryliberal. The account seems to be a sock puppet for 4437.

**musings033 (forum guest):** I agree with triune's assessment on the economics of physical goods (also keep in mind that there is actually an infinitesimal chance of possessing something that another form of intelligence would want). As for specific ideas to transmit, why not just try to get to know each other first? The best way to do that, I think, is to be open and honest about our world, in hopes of reciprocity. We can compile a narrative — a BBC 'documentary' if you will — on Earth, our lands and oceans and all our beautiful organisms and ideas and constructions, culminating with the growth and maturation of a baby. Meanwhile, we'll also send a documentary that details our killing fields and memorials, our faces of fear and courage, our acts of barbarism and compassion upon one another and upon our planet. I don't know how we'll show them everything, but if the beings cannot see, we will make them hear. If they cannot hear, we will make them feel. We are good and beautiful, and we have given Earth scars that will never properly heal. And so we will present to Gliese 581 our humanity, the purest and the vilest thing in our possession. That, in itself, is the first idea worth trading.

**twinkle\_twinkle:** That's a very complicated 'first idea'.

**triune:** I fully support a gift to Gliese 581. The only thing I'd warn against is miscommunication. Should we even make it seem as if we expect anything in return?

**thisisfutile888 (forum guest):** Okay, musings, I think you've logged into the wrong forum. Space Cakes Conversations is another Google button away. Seriously, 'documentaries' about a human baby and another centred around humanity's thirst for blood and conquest? The first idea makes me want to vomit, and the second will get us all killed. Why not load up *Free Willy* and *The Little Mermaid* and force a couple of death-row prisoners to care for them? Gliese 581 will piss delight.

**twinkle\_twinkle:** I've banned thisisfutile888. The future is just lost on some people. ■

**Ken Liu** is a lawyer and programmer. You can read more of his fiction at <http://kenliu.name/stories>. **Shelly Li** does not believe in working, although her first novel, *The Royal Hunter*, is forthcoming from Penguin Books in 2011. To learn more, visit [www.shelly-li.com](http://www.shelly-li.com).